日英語韻律特徴と発音学習

18530517

平成18年度～平成19年度科学研究費補助金

（基盤研究（C））研究成果報告書

平成20年4月

研究代表者　山田　純

広島大学大学院総合科学研究科教授

はしがき

本研究は、日本人の英語発音を改善するための基礎研究である。第１章は、本研究の中核をなす。日本人大学院生１名を被験者として、英語テキストの音読を録音し、韻律特徴の記述を試みた。その結果、英語音読のピッチ幅が著しく小さく単調である、新旧情報とピッチ高低に相関が認められない、名詞句は、高低型および平板型の２種類に限定される、代名詞と名詞（“he” and “Ben”）では相対的に前者のピッチが高い、等が明らかになった。これらは、日本語韻律の干渉と考えられ、日本語の音読特徴を明らかにする必要がある。

第２章では、日本語の範読の分析をおこなった。範読として、CD に録音されている専門家（俳優）の音読における文節ごとのピッチ、その自己相関、格助詞「は」の持続時間などを測定し、日本語の発話特徴を記述した。

第３章は、日本人大学生を被験者とした。日本人が英語を音読する場合、鼻音率が低いように感じられるが、実際の計測データはないようである。そこで、日英語の短い文章および句の音読をさせ､その鼻音率をナゾメーターで計測した。その結果、低圧の口音には２つのタイプが、また被験者には３つのタイプが存在することが示唆された。

第４章は、日本人吃音者の日本語音読について、聞き手あり条件と聞き手なし条件で音読速度、吃音率、ピッチがどのように影響を受けるかについて、また吃音が発生しそうに感じた場合に､「えー」というつなぎの語を発することの効果について検討を加えた。結果は、聞き手なし条件で吃音率が下がり、ピッチも約５Hz 下がった。ピッチについては、これまでこのような結果は、先行研究では報告されていない。

第５章では、第４章の結果を受けて、A-B-A-B パラダイムを用いて、聞き手の有無の効果をより厳密に検討することを試みた。聞き手なしのＢ条件では、平均の基本周波数が聞き手ありのＡ条件よりも有意に低い結果を得て、聞き手がおよぼす否定的効果が実証された。これがＢ条件における音読の音響・音声特徴にど

のような効果をもたらしているかは今後の課題となる。

このほか、本報告書に記載するにいたらなかったが、いくつかの分析が進行中もしくは構想中である。第5章の関連では、健常者の日英語音読について、逆に聞き手あり条件で、基本周波数が下がる傾向を得ている。実験条件でA-BまたはB-Aパラダイムを用いたため、比較ができにくいが、さらなる実験が必要となる。また、英語名詞句に限定し、構成素の形容詞と名詞の基本周波数の測定が進行中で、英語範読ではさまざまなパタンが観察されつつある。これらは、展開研究として継続予定である。

研究組織

研究代表者：　山田　純　（広島大学大学院総合科学研究科教授）

交付決定額

	直接経費	間接経費	合計
平成18年度	1,000,000	0	1,000,000
平成19年度	600,000	180,000	780,000
総計	1,600,000	180,000	1,780,000

研究発表

(1) 雑誌論文

Yamada, J. and Homma, T. (2007). A simple treatment for stuttering: The effect of audience and repetition. *Medical Hypotheses*, 69, 1196-1199.

Contents

Chapter 1

Intonational Interference from L1:

The Case of a Japanese Student in Oral Reading in English

Co-authored with Harumi Ototake

Abstract

This case study examined a Japanese student's oral reading performance in English and revealed four characteristic features of intonation patterns: (1) the pitch range at the sentence and phrase levels was much narrower for the student than for the native English speakers, (2) unlike the native speakers, the student exhibited no correlation between pitch and new/old information, (3) unlike the native speakers, he exhibited only two pitch patterns (a high-low and a monotone pattern) at the noun phrase level, and (4) his pitch was relatively higher on "he" than "Ben," as compared with the native English speakers'. These features are attributable at least in part to interference from his native language phonological and grammatical systems. Implications 'and related issues arising from these findings are presented.

Keywords: L1 intonational interference; Japanese student; Pitch range; New/old information; Paucity of pitch pattern; Pronunciation of *he*

1. Introduction

Chun (2002) states that "[intonation] is easy for adults to maintain and retain in the L1, yet difficult, if not impossible, for adults to learn in an L2" (p. xiii). This implies that, where there are significant intonational differences between two languages, interference from L1 to L2 is inevitable. But the question of specifically how interference takes place remains almost unknown. The present case study explores this question by examining the English oral reading performance of a Japanese student.

1.1. Previous studies

To our knowledge, there are only three studies on L1 intonational interference in Japanese students of English (Loveday, 1981; Wennerstrom, 1994, and Sugito, 1996), and their findings motivate further research. Loveday (1981) found that Japanese males ($N = 3$) on an English oral reading task transferred their narrower Japanese pitch ranges to English although that did not seem to be the case for Japanese females ($N = 2$). This is the first example of L1 intonational interference found for Japanese students of English. However, because what Loveday dealt with was only three speech formulas, "oh hello" (*aa konnichiwa*), "thank you" (*arigato gozaimasu*), and "bye" (*sayonara*), the generality of this finding was questionable.

Wennerstrom (1994) compared F_0 values of some key words in the text read by Japanese learners of English (five males and five females) and native English speakers, and did not find gross deviation from the English norm among the Japanese students. She noted, however, that the pitch range was smaller for some phrases in the Japanese group than in the native English speaker group. Thus, for example, the mean F_0 values for "hard" and "rain" in the phrase "in a hard rain" were, respectively, 177 Hz and 150 Hz for the

Japanese subjects and 184 Hz and 141 Hz for the English speakers. Wennerstrom did not discuss the possibility that the narrower pitch range of the Japanese group could be a consequence of interference from the Japanese pitch system.

Sugito (1996) had six Japanese speakers and two English speakers read a short story in English (an English translation of an old Japanese tale) and examined the effect of new/old information of one particular word, "peach," on pitch contour. The word appeared three times, and the mean F_0 values were 199 Hz (*SD* = 26), 246 Hz (*SD* = 26), and 264 Hz (*SD* = 34) on the first, second, and third occurrences for the Japanese subjects, and approximately 400 Hz, 100Hz, and 180 Hz for one English speaker, and 320 Hz, 180 Hz, and 200 Hz for the other. Sugito suggested that these differences arose from the possibility that every Japanese word has an inherent pitch pattern which may be less likely to be influenced by syntactic, semantic, and pragmatic functions, as compared with English (cf. Abe, 1955). This finding can be taken as constituting a second example of L1 interference.

To date there have been just the two examples, cited above, of L1 intonational interference with which Japanese students of English seem to exhibit. However, these findings are not free from methodological problems. The text used in each study, for example, was very short: 110 words in Loveday (1981), 114 words in Wennerstrom (1994), and 78 words in Sugito (1996). (See Henton, 1995, for other criticisms of Loveday, some of which also apply to Wennerstrom, 1994, and Sugito, 1996). We need to verify the reliability of these findings. In addition to a reconsideration of these findings, we attempt to examine two more probable cases for L1 intonational interference in this study.

1.2. Narrow pitch range

A key aspect of the Japanese intonational system may largely be represented by the inherent pitch accent for each lexical item in Japanese (e.g., Beckman, 1996; Beckman & Pierrehumbert, 1986). Beckman and Pierrehumbert (1986), for example, observe that "Japanese also has lexically unaccented words and consequently can have well-formed utterances without any pitch accents, which would be impossible in the English intonation system" (pp. 305-6). Thus, the fact that monotonous sentences with zero accents can be acceptable in Japanese can be a crucial difference between English and Japanese which causes intonational interference for Japanese students of English.

However, in looking over pitch contours of noun phrases in previous studies on Japanese intonation (e.g., Gussenhoven, 2004; Pierrehumbert & Beckman, 1988; Poser, 1984; Selkirk & Tateishi, 1991; Venditti, 2006), we find that some pitch ranges are not always small. For the adjective-noun phrase, "uma'i mame'-wa" (*good-tasting beans*) in Pierrehumbert & Beckman (1988, Figure 4.7), for example, the phrasal high in the accented adjective "uma'i" bears the culminative tonal prominence, with the F_0 value rising from about 125 Hz to 260 Hz. The F_0 value then desends to the level of 150 Hz on the following noun "mame'-wa."

On the other hand, pitch ranges for English utterances are not always large, either. Take some findings from Cooper and Sorensen (1981, pp. 30-33), for example, where native English speakers (five males and five females) were given several sentences on an oral reading task. For the sentences, (1) "The *cat* was asleep in the *tree*" and (2) "The *cat* that Sally owned was asleep on the large branch in the *tree*," the mean F_0 peak values were respectively 195 Hz (*SD* = 65) and 156 Hz (*SD* = 60) for "cat" and "tree" in the first sentence, and 206 Hz (*SD* = 76) and 156 Hz (*SD* = 62) in the second sentence. Of course, the size of pitch range varies from speaker to speaker and from condition to condition (e.g., Bolinger, 1986; Umeda, 1982; Levis & Pickering, 2004).

1.3. Pitch and information structure

We have a plethora of English studies on the relationship between new/old information, focus, and contrast on the one hand, and pitch and stress on the other (see, e.g., Bolinger, 1986; Breul, 2004; Chomsky, 1971; Cooper and Sorensen, 1981; Cooper, Eady, & Mueller, 1985). In the case of Japanese, there are many studies on the nature of new/old information, focus, and contrast (e.g., Kuno, 1978), but their relations to pitch have not been intensively investigated. In Japanese, new/old information, focus, and contrast are marked by means of word order, case particles (e.g., "–wa" and "–ga"), stress, and syllable lengthening (e.g., Aizawa, 1981; Kuno, 1978), and these factors might tend to make the true effect of pitch on information structure (new/old information) difficult to observe. Whether pitch plays no role at all, as suggested by Sugito (1996), must thus be confirmed examining some more data.

This question may not be easy to answer, because there are possible confounding factors. For example, if the F_0 value is high for a sentence-initial noun with new information, we cannot determine if the high value is due to sentence position, new information, or both. If, however, we find no significant correlation between pitch and information structure (new/old information), it would be reasonable to conclude that the effect of information structure on pitch is not significant. (If there is a significant correlation, we would have further to examine the linguistic contexts around the target words.)

1.4. A third example of L1 intonational interference

In search of a third example of L1 intonational interference, we consider the paucity of intonation contours in Japanese which may be associated with narrow pitch range. Beckman and Pierrehumbert (1986) assert that, largely due to the nature of the lexical accent system in Japanese, "the range of possible intonation variation is considerably smaller than in the English intonation system" (p. 306). According to them, Japanese has only one pitch contour, a H*+L (High*+Low) shape for a phrase aside from a boundary low tone, L%, and a phrasal peak H tone, whereas English has six, H*, L*, H*+L, H+L*, and L+H*, where the asterisk designates a metrically strong tone (see Ladd & Schepman, 2003, for criticisms of the distinction between H* and L+H*). This paucity of intonation contours in Japanese may lead to the paucity of intonation contours in the Japanese student's oral reading in English within the narrow range.

1.5. A fourth example of L1 intonational interference

As a fourth probable example, we investigate the distribution of pitches of the personal pronoun "he." This pronoun (and some other pronouns) can be of interest in two respects. First is the pronunciation of "he," i.e., /hiː/ or /hi/, which many Japanese students, whether at the beginning or advanced level, transform into /hii/. All else being equal, this "heavy" syllable word may readily attract accent (compare, e.g., "boot*y*" and "boot*ie*" in English). Second is the special linguistic status of "kare," the Japanese counterpart of "he," where the first mora has a high tone and the second one a low tone. The word "kare" is defined as a third singular personal pronoun in Japanese dictionaries, but it functions just as a common noun, as the following examples illustrate: "watashi-no kare" (**my he*), "kare shi" (**Mr. he*), and "kare to Suzuki san no kenkyu" (**he and Mr. Suzuki's research*). This implies that because "kare" has an inherent pitch accent on its first mora, it may receive accent in an utterance in just the same way as other common nouns. Then if we assume that this aspect of "kare" transfers to the English pronoun "he" when a Japanese student reads an English passage, we expect that the distribution of F_0 values of "he" in the Japanese student is similar to that of other personal nouns, e.g., "Ben," which would be in marked contrast with the case of native English speakers. We will verify this possibility below.

2. Method

2.1. Subjects

The target subject was JH, a Japanese male student age 24, who was studying the work of Middle English grammarians in a master's program at a state university in Japan. JH was a speaker of Aichi Japanese, spoken around Aichi Prefecture, located approximately in the middle of Honshu, the main island of Japan. This dialect is similar to Tokyo Japanese. He had not lived in an English-speaking country, but given his 12-year experience of learning English as an L2, his overall proficiency in English was considered to be at an advanced level. His English pronunciation, however, seemed to have reached a plateau at an intermediate level.

Two native speakers of English were employed as subject controls. One was a Californian male age 51, a phonology professor at a Japanese state university. The other was Jack Moyles, a professional narrator who also served as news caster in radio and television for Voice of America in

Washington, D.C. A commercially available cassette tape on which Moyles' oral reading was recorded was utilized (see below). We will call them NS1 and NS2, respectively.

We had two Japanese actors, Minori Terada and Toru Emori, as subjects. We used their oral readings recorded on commercially available CDs. The actors were respectively 60 years old and 58 years old when their CDs came out from Shincho-sha, Tokyo, in 2002. Both were native speakers of Tokyo Japanese. We will call them JP1 and JP2, respectively.

2.2. Materials

The reading material was an English story entitled "Case of the Contraband Camera," which was taken from Lipman (1970/1990, pp. 4-6), a textbook for English language classes at college freshman levels. The story consisted of 11 paragraphs, 42 sentences, and 650 words (see Appendix for the first three paragraphs).

This story was translated into Japanese by the first author. The Japanese translation consisted of 11 paragraphs, 52 sentences, and 330 phonological words. The number of characters composing the text was 1,550. Note that the English text does not necessarily correspond to the Japanese translation in a word-to-word or sentence-to-sentence manner, so that the number of tokens of "Ben" in the English text, for example, was not necessarily the same as that of "Ben" in the Japanese translation.

The Japanese text read by JP1 was the first part of the essay entitled "Lemon," written by Motojiro Kajii (1901-1932) in 1931. The number of sentences was 41. (The essay describes how the author, depressed by an unknown, evil mass in his mind, bought a lemon at a cozy fruit store, and then left it in a bookstore which he used to love to visit. Because the events and scenes described there are, in general, low-key, and less emotional, the effects of emotion on speech rate do not seem great.) The other text read by JP2 was a Japanese translation of a simplified version of F. M. Dostoevsky's *Crime and Punishment*, and the first 41 sentences were used as test material.

2.3. Procedure

JH and NS1 were tested individually in a quiet room. Both read the text silently before the oral reading test session. JH was given the English text first on Day 1 and then the Japanese translation on Day 2. The recordings of JH and NS1 were made via a microphone (Shure Model SM48) connected to a *Computerized Speech Lab*, Model 4500 (*CSL*, Kay Elemtrics) with a sampling rate of 20 kHz and a 16-bit resolution.

2.4. Measurements and Analyses

Measurements of JH and NS1's oral reading were made using a *CSL*. NS2's taped material was analyzed by the *CSL*. The two Japanese actors' CD recorded materials were analyzed by an *Animo SUGI speech analyzer* (Version 1.0.7.8), the sampling rate being 44 kHz with a 16-bit resolution. In measuring and analyzing pitch, we followed Ladefoged's (2003, pp. 75-90) suggestions. The target items for measurements included content words. We measured 329 content words and some function words for the English text, and 330 *bunsetsu* (a noun plus any following postpositions) for the Japanese translation. We used a similar procedure in measuring and analyzing the Japanese actors' reading performances.

Following Cooper and Sorensen (1981) and others, we measured the F_0 peak for each content word (but cf. 't Hart, Collier, & Cohen, 1990, for some possible drawbacks of this method). The pitch range for each sentence was defined as the max F_0 peak minus the min F_0 peak of the sentence. The same definition was used for the pitch range for noun phrases.

In oral reading in Japanese, some words were found to be at falsetto, their F_0 values going up to as high as 400 Hz or more. Such words were excluded from analysis. Also excluded were some sentence-final (auxiliary) verbs which faded into voicelessness.

3. Results

3.1. Narrower Pitch Range

3.1.1. Pitch range at the sentence level

The mean max and min peak F_0 values of English content words and Japanese phonological words in each sentence are presented in Table 1. (Semitones also characterize subjects' pitch ranges: 4.4 ST, 10.2 ST, 12.2 ST, 4.1 ST, 8.5 ST, and 8.8 ST, for JH in English, NS1, NS2, JH in Japanese, JP1, and JP2, respectively. However, because results essentially remain the same, we confine ourselves to the Hertz in this report.) Regarding pitch range in English reading, a one-way ANOVA indicated that the effect of subject (JH, NS1, and NS2) was significant, $F(2, 41) = 99.19$, $p < .001$. JH's mean pitch range (38 Hz) was significantly smaller than those of the two native English speakers (97 Hz and 110 Hz), $t(41) = 10.38$, $p < .001$, and $t(41) = 14.58$, $p < .001$.

[TABLE 1 GOES NEAR HERE]

JH's narrow pitch range could be ascribed more to his lower max F_0 values than to his min F_0 values. For max F_0 values, the effect of subject was highly significant, $F(2, 41) = 53.46$, $p < .001$, and JH's mean min F_0 value was much lower than the native speakers', the difference being 52 Hz between JH and NS1, $t(41) = 9.20$, $p < .001$, and 50 Hz between JH and NS2, $t(41) = 10.00$, $p < .001$. For min F_0 values, the effect of subject was also highly significant,

$F(2, 41) = 64.05$, $p < .001$, but JH's mean min F_0 value was higher than the native speakers', the difference being 8 Hz between JH and NS1, $t(41) = 3.83$, $p < .001$, and 22 Hz between JH and NS2, $t(41) = 12.30$, $p < .001$.

JH's mean pitch range in the English text (38 Hz) was significantly higher than his mean in the Japanese translation (33 Hz), $t(41) = 2.42$, $p < .05$. As for Japanese reading, a one-way ANOVA showed that the effect of subject (JH, JP1, and JP2) was significant, $F(2, 121) = 50.13$, $p < .001$. JH's mean pitch range (33 Hz) was significantly smaller than those of the Japanese actors (54 Hz and 67 Hz), $t(81) = 4.84$, $p < .001$, and $t(81) = 5.38$, $p < .001$.

The effect of language (English and Japanese) in L1 reading on pitch range was also significant. That is, the pitch range was higher in English than in Japanese; for example, the mean pitch range for NS1 (i.e., 97 Hz, the lower one between the two native speakers) was significantly higher than that for JP2 (i.e., 67 Hz, the higher one between the two Japanese actors), $t(81) = 3.79$, $p < .001$.

3.1.2. Pitch range at the phrase level

Essentially similar results were observed at the phrasal level. Table 2 presents the mean peak F_0 values of the adjective (i.e., F_0(Adj)), noun (F_0(Noun)), and absolute difference between the peak F_0 values of the adjective and noun ($|F_0(Adj) - F_0(Noun)|$) in each adjective-noun phrase.

[TABLE 2 GOES NEAR HERE]

The mean F_0(Adj) and F_0(Noun) values are less informative because F_0(Adj) was sometimes higher, and sometimes lower than F_0(Noun) in each subject, and so we paid attention to $|F_0(Adj) - F_0(Noun)|$ and pitch ranges indexed by standard deviations. For the values of $|F_0(Adj) - F_0(Noun)|$, the main effect of subject (JH, NS1, and NS2) was significant, $F(2, 50) = 8.88$, p

< .001. The mean $|F_0(\text{Adj}) - F_0(\text{Noun})|$ was significantly smaller for JH than both for NS1, $t(25) = 4.01$, $p < .001$, and for NS2, $t(25) = 3.96$, $p < .001$.

A similar pattern of results was also observed for the noun-noun phrases, although the number of items was small (Table 3).

[TABLE 3 GOES NEAR HERE]

3.2. Pitch and New/Old Information

We selected only two words for analysis. The first was "camera," which was one of the key words, appearing nine times in this story. The context in which the word appeared, and the mean F_0 values are shown in Table 4. Here too, we see that JH's pitch range is very narrow, which itself may blur the distinction between new and old information.

[TABLE 4 GOES NEAR HERE]

In Table 4 the nine tokens are divided into two groups with respect to new/focused and old information. Six of them carried new/focused information, and the remaining three carried old information. The mean F_0 values of the words with new/focused and old information were 136 Hz and 141 Hz, 173 Hz and 138 Hz, and 165 Hz and 141, for JH, NS1, and NS2, respectively. The chief difference between JH and the two native English speakers was found in the distribution of new-information-bearing words. That is, the mean F_0 value of those words tended to be lower for JH than for the two native English speakers, $t(5) = 2.47$, $p < .06$, and $t(5) = 1.67$, *n.s.*, respectively, which could also be taken as another manifestation of JH's narrow pitch range.

The second word we selected was another key word, "Ben," which appeared 14 times. The results were essentially the same as those for "camera." The mean F_0 values were 155 Hz ($SD = 17$), 166 Hz ($SD = 45$), and

179 Hz (*SD* = 51) for JH, NS1, and NS2, respectively. The mean F_0 values of the words with new/focused (*N* = 8) and old information (*N* = 6) were 155 Hz (*SD* 16) and 154 Hz (*SD* = 18), 181 Hz (*SD* = 55) and 154 Hz (*SD* = 37), and 227 Hz (*SD* = 32) and 144 Hz (*SD* = 27), for JH, NS1, and NS2, respectively.

3.3. Paucity of Pitch Patterns

Let us return to the results concerning F_0(Adj), F_0(Noun), and $|F_0(\text{Adj}) - F_0(\text{Noun})|$ in Table 2. We here consider the correlations between F_0(Adj), F_0(Noun), and $|F_0(\text{Adj}) - F_0(\text{Noun})|$ for each subject in the English condition. For JH, the correlation was .57, $p < .01$, between F_0(Adj) and F_0(Noun), .33, $.05 < p < .10$, between F_0(Adj) and $|F_0(\text{Adj}) - F_0(\text{Noun})|$, and -.58, $p < .01$, between F_0(Noun) and $|F_0(\text{Adj}) - F_0(\text{Noun})|$; for NS1, it was -.07, *n.s.*, .72, $p < .001$, and -.75, $p < .001$, respectively; and for NS2, it was .17, *n.s.*, 77, $p < .001$, and -.50, $p <.01$, respectively.

Overall, these findings suggest that the pitch contours of the adjectives and nouns in JH were impoverished as compared with those for the native speakers. In fact, inspection of individual phrases suggests that JH had two patterns in the narrow pitch range: a flat pattern and a high-low pattern. Let us arbitrarily use 10 Hz as a significant change in pitch; that is, if the differences between F_0(Adj) and F_0(Noun) were 10 or less than 10 Hz we assume that the adjective-noun phrase has a flat pattern. Given this criterion, 69% (*N* = 18) of JH's adjective-noun pairs were flat, and the remaining 31% (*N* = 8) were a high-low shape, the latter of which is the typical shape of Japanese noun phrases (Beckman & Pierrehambert, 1986). In the case of NS1, only 11% (*N* = 3) were flat, 56% (*N* = 15) had a high-low shape, and 33% (*N* = 9) had a low-high shape. Similarly for NS2, 19% (N = 5) were flat, 44% (*N* = 12) had a high-low shape, and 37% (*N* = 10) had a low-high shape. Figure 1 presents an example which illustrates a clear difference between JH and the native speakers.

[FIGURE 1 GOES NEAR HERE]

Similar patterns of F_0 distributions were obtained for the N1-N2 phrases (see Table 3). For JH, the F_0(N1) values were significantly higher than the F_0(N2) values, $t(5) = 3.93$, $p < .05$, but two of the six phrases have a flatter shape in terms of the above criterion (see Figure 2). In JH's Japanese, the difference in F_0 between N1 and N2 was small, $t(6) = 0.47$; five of the seven noun-noun phrases were of flatter type. On the other hand, results were not consistent between the two native speakers, i.e., for NE1, $t(5) = 0.71$, *n.s.*, and for NE2, $t(5) = 2.79$, $p < .05$.

[FIGURE 2 GOES NEAR HERE]

3.4. Personal Pronoun and Noun in English and Japanese

As stated in the introduction, the Japanese personal pronoun "kare (he)" behaves like personal nouns in terms of phonology and syntax. This section verifies the hypothesis that JH's F_0 values for "he," "Ben," "kare," and "Ben" (in Japanese) are all comparable, while English speakers' F_0 values are lower for "he" than for "Ben." The mean F_0 values for those words are given in Table 5.

[TABLE 5 GOES NEAR HERE]

By and large, the results were consistent with the hypothesis. A 2×2 (Language: English and Japanese, and Word Category: personal pronoun and personal noun) ANOVA on JH's responses indicated that none of the effects were significant, $F(1, 51) = 0.39$, for language, $F(1, 51) = 0.20$, for word category, and $F(1, 51) = 0.44$, for the interaction between language and word category. As shown in Table 5, the mean F_0 values were not significantly different between "he" and "Ben" in English on the one hand, and between "kare" and "Ben" in Japanese on the other. Neither were those between "he" and "kare," $t(x) = 0.96$. In contrast, the mean F_0 value of "he," not surprisingly, tended to be significantly lower than that of "Ben" for NS1 and NS2.

4. Discussion

4.1. Major Findings

This case study gave JH, a Japanese student, an English oral reading task, and revealed that (1) the pitch range at the sentence and phrase levels was much narrower for the student than for native English speaker (Tables 1 to 3), (2) the student exhibited no correlation between pitch and new/old information (Table 4), (3) he exhibited only two pitch patterns (a high-low and a monotone pattern) at the noun phrase level, and (4) his pitch was relatively higher on "he" than "Ben," as compared with the native English speakers' (Table 5). The first finding is taken to generalize Loveday's (1981) finding concerning Japanese males' narrower pitch ranges in reading English formulaic sentences. The second one substantiates Sugito's (1996) finding. The third and fourth were newly discovered in this study.

While all these findings can at least partially be accounted for as consequences of L1 interference, they, except for the last one, may be put together under the rubric of monotonous patterns. Actually, the third finding that many English noun phrases in JH's were flat or their pitch ranges were compressed is directly related to the first finding that JH's pitch ranges at the phrase level were narrow. And both are considered due to the effect of Japanese noun phrases with no pitch accents. Even for Japanese noun phrases having a high-low pattern, the pitch range was small. That can thus be taken as contributing to the narrow pitch range in JH's English phrases. As to no effect of information structure on pitch, Japanese noun phrases with no accents can also play a role to blur the distinction between new/focused information and old information in JH's English pitch contours.

The fourth example of L1 intonational interference was the F_0 for the personal pronoun "he" which was even higher, though not significant, than the personal noun "Ben" in JH (Table 5). As suggested in the Introduction, this unique status of "he" in JH can be attributable to its pronunciation /hii/ and/or its Japanese counterpart "kare," which functions like personal nouns such as "Ben." We can't determine which is more influential, the pronunciation /hii/ or the Japanese "kare" in this study. But we prefer the latter, and suggest that "he" is pronounced emphatically as /hii/ as a consequence of the effect of "kare." The fact that it is very unlikely that "the" of "the average hobbyist" and others becomes /ðiː/ in JH seems consistent with this interpretation.

4.2. Generality of the Findings and Implications for Further Research

The generality of the findings of a single case study is limited in many ways. What we have found in this study is only the tip of the iceberg. We will here briefly discuss some of many remaining problems centering around the narrow pitch range in JH's oral reading in English and Japanese: individual differences, language mode, and educational implications.

First, we are well aware that the present findings are not consistent with some data from other studies. In addition to the examples presented in the Introduction, we mention two more studies here. Selkirk and Tateishi (1991) measured F_0 values of nouns in noun phrases read by two male (and two female) speakers of Tokyo Japanese. One would immediately notice that in their study the pitch ranges greatly differed between these two males, one being moderately large and the other being very large. For example, in the sentence, "[[Ao'yama-no] {[[Yama'guchi-ga] ani'yome-o] yond} (Mr. Yamaguchi from Aoyama called his sister-in-law)," whose structure may be represented by [[N1-no]N2-ga]{[N3-o]Verb}, the mean F_0 values of N1 and N2 were 200 Hz and 147 Hz for one speaker, and 204 Hz and 112 Hz for the other. In a similar sentence, "[urite-wa]{[[sho'hin-no] shoyu'sha-de] aru} (The seller owns the merchandise)" with the structure [N1-wa]{[[N2-no]N3-de]Verb}, which appeared in our Japanese translation, the F_0 values of N2 and N3 were 162 Hz and 141 Hz for JH.

Wennerstrom (2001), on the other hand, investigated the distribution of pitch contours in English narratives collected from six English speakers in a graduate seminar and six Japanese students in an English conversation class. The purpose of this research was not to compare pitch ranges, but we find that one Japanese male student's pitch ranged from 85 to 197 Hz (pitch range = 112 Hz). Female students' pitch ranges were still greater.

Apart from some methodological differences, these findings suggest that individual differences may be involved. The differences in the findings between the Selkirk and Tateishi study and the present one may indicate that in terms of pitch range in Japanese, there are individual differences even among males. The differences in the finding between Wennerstrom (2001) and the present study may imply individual differences in L2 ability and/or language modes (spontaneous speech versus oral reading).

Also related to this issue is the tendency that the professional narrator or readers, NS2, JP1, and JP2, had larger pitch ranges than did "lay" persons, NS1 and JH (see Tables 1 to 3). Those professional readers might have acquired the ability to exaggerate accented and unaccented parts in utterances to more vividly convey the semantically, pragmatically, or emotionally important message therein. By combining this professionalism with language effect, the six subjects (including two JHs, one in the English condition and the other in the Japanese) are ordered, from large to small, in terms of the size of pitch range: NS2, NS1, JP2, JP1, JH in English, and JH in Japanese. JH's larger, though not significant, pitch range in English than in Japanese may be interpreted as suggesting an effect of English intonation learning.

Another possibility which may account for some apparent inconsistencies across studies is that JH and others may use two or more sets of pitch range in different contexts and/or according to different language modes, e.g., a narrower range in oral reading and a wider range in spontaneous speech. Consistent with this interpretation is the case of DC, a native speaker of English reported by Umeda (1997), which showed different styles in reading a text and talking in passive and active conversations; that is, peak F_0's of phrases showed a small range (45 Hz) in the text reading, a moderate range (72 Hz) in the passive conversation, and a large range (121 Hz) in the active conversation. DC's case may apply to JH and other Japanese students.

Finally, there are many educational implications. To improve his pronunciation skill in English, JH would have a lot of things to try. Among others, JH would have to enlarge his pitch range in oral reading in English, specifically heighten his max F_0 peaks for noun phrases. Second, he would have to try to enrich pitch patterns; in particular, if we assume that the classification of pitch patterns proposed by Beckman and Pierrehumbert (1986) and Pierrehumbert and Beckman (1988) is valid, he would have to learn six pitch accent shapes, H*, L*, H*+L, H+L*, and L+H*. However, because these are only surface patterns, he would have to learn a set of ordered constraints which allow the optimal pattern for a given phrase (e.g., Gussenhoven, 2004). For example, although JH seems to always abide by the compound rule which stipulates that the right-hand constituent of compound words is unaccented (e.g., Chomsky & Halle, 1968, p. 18; Gussenhoven, 2004, p. 277), he would have to learn under what condition the rank of this rule is lowered. On the other hand, researchers have to investigate the learnability of such constraints in first and second language learners (e.g., Cruttenden, 1981; Snow, 1998). All these challenges are left to future research.

References

Abe, I., 1955. Intonational patterns of English and Japanese. Word, 11, 386-398. Reprinted in D. Bolinger (1972). (Ed.). Intonation, pp. 337-347. London: Harmondaworth.

Beckman, M. E., 1996. The parsing of prosody. Language and Cognitive Processes, 11, 17-67.

Beckman, M. E. and Pierrehumbert, J. B., 1986. Intonational structure in Japanese and English. Phonology Yearbook, 3, 255-309.

Bolinger, D., 1986. Intonation and its parts: Melody in spoken English. London: Edward Arnold.

Breul, C., 2004. Focus structure in generative grammar: An integrated syntactic, semantic and intonational approach. Amsterdam: John Benjamins.

Chomsky, N., 1971. Deep structure, surface structure, and semantic interpretation. In D. D. Steinberg and L. A. Jakobovits (Eds.). Semantics: An interdisciplinary reader in philosophy, linguistics and psychology, pp.183-216. Cambridge: Cambridge University Press.

Chomsky, N. and Halle, M., 1968. The sound pattern of English. New York: Harper & Row.

Chun, D. M., 2002. Discourse intonation in L2: From theory and research to practice. Amsterdam: John Benjamins.

Cooper, W. E. and Sorensen, J. M., 1981. Fundamental frequency in sentence production. New York: Springer-Verlag.

Cooper, W. E., Eady, S. J., and Mueller, P. R., 1985. Acoustic aspects of contrastive stress in question-answer contexts. Journal of the Acoustical Society of America, 77, 2142-2156.

Cruttenden, A., 1981. Falls and rises: Meanings and universals. Journal of Linguistics, 17, 77-92.

Gussenhoven, C, 2004. The phonology of tone and intonation. Cambridge: Cambridge University Press.

't Hart, J., Collier, R. & Cohen, A, 1990. A perceptual study of intonation: An experimental-phonetic approach to speech. Cambridge: Cambridge University Press.

Henton, C, 1995. Pitch dynamism in female and male speech. Language & Communication, 15, 43-61.

Kuno, S, 1978. Danwa no bunpo [A grammar of discourse]. Tokyo: Taishu-kan.

Ladd, D. R. and Schepman, A, 2003. "Sagging transitions" between high pitch accents in English: Experimental evidence. Journal of Phonetics, 31, 81-112.

Ladeforged, P, 2003. Phonetic data analysis: An introduction to fieldwork and instrumental techniques. Oxford: Blackwell Publishing.

Levis, J. and Pickering, L, 2004. Teaching intonation in discourse using speech visualization technology, System, 32, 505-524.

Lipman, M., 1970/1990. Point of law. Vol. 2. Tokyo: Nan'un-do.

Loveday, L, 1981. Pitch, politeness and sexual role: An Exploratory investigation into the pitch correlates of English and Japanese politeness formulae. Language and Speech, 24, 71-89.

Pierrehumbert, J. B. and Beckman, M. E, 1988. Japanese tone structure. Cambridge, Mass.: MIT Press.

Poser, W. J, 1984. The phonetics and phonology of tone and intonation in Japanese. PhD dissertation, MIT.

Selkirk, E. and Tateishi, K, 1991. Syntax and downstep in Japanese. In C. Georgopoulos and R. Ishihara (eds.). Interdisciplinary approaches to language: Essays in honor of S. -Y. Kuroda, pp. 519-543. Dordrecht: Kluwer.

Snow, D., 1998. Children's imitations of intonation contours: Are rising tones more difficult than falling tones? Journal of Speech, Language, and Hearing Research, 41, 576-587.

Sugito, S., 1996. Nihonjin no eigo [Japanese speakers' English]. Tokyo: Izumi Shoin.

Umeda, N., 1981. F0 declination is situation dependent. Journal of Phonetics, 10, 279-291.

Umeda, N., 1997. Fundamental frequency rule for English discourse. In S. Kiritani, Hirose, H., and Fujisaki, H. (Eds.). Speech production and language: In honor of Osamu Fujimura, pp. 131-156. Berlin: Mouton de Gruyter.

Venditti, J. J., 2006. Prosody in sentence processing. In M. Nakayama, Mazuka, R., and Shirai, Y. (eds.). The handbook of East Asian psycholinguistics. Vol. 2, pp. 208-217. Cambridge: Cambridge University Press.

Wennerstrom, A., 1994. Intonational meaning in English discourse: A study of non-native speakers. Applied Linguistics, 15,399-420.

Wennerstrom, A., 2001. Intonation and evaluation in oral narratives. Journal of Pragmatics, 33, 1183-1206.

Table 1

Mean Max and Min F_0 Peaks for Sentences

		Max F_0	Min F_0	Difference
English text				
JH	*N*	42	42	42
	Mean	168	130	38
	SD	11	5	12
NS1	*N*	42	42	42
	Mean	220	122	97
	SD	35	12	35
NS2	*N*	42	42	42
	Mean	218	108	110
	SD	36	11	35
Japanese translation				
JH	*N*	42	42	42
	Mean	156	123	33
	SD	9	10	13
Other Japanese texts				
JP1	*N*	41	41	41
	Mean	139	85	54
	SD	26	9	26
JP2	*N*	41	41	41
	Mean	168	101	67
	SD	38	23	38

Table 2

Mean Peak F_0(Adj), F_0(Noun), and F_0(Dif) Values for Adj+Noun Phrases

		Adjective	Noun	Dif[1]
English				
JH	*N*	26	27	26
	Mean	148	140	9
	SD	10	11	9
NS1	*N*	27	27	27
	Mean	156	150	36
	SD	31	33	30
NS2	*N*	27	27	27
	Mean	161	148	40
	SD	48	35	39
Japanese translation				
JH	*N*	24	25	22
	Mean	149	139	14
	SD	9	13	13
Other Japanese texts				
JP1	*N*	14	14	14
	Mean	124	108	18
	SD	20	21	16
JP2	*N*	13	13	13
	Mean	156	131	27
	SD	38	46	12

Note. 1) $\text{Dif} = (1/N)\sum | F_0(\text{Adj}) - F_0(\text{Noun}) |$.

Table 3

Mean F_0(N1), F_0(N2), and F_0(Dif) Values for N1+N2 Phrases

		Noun 1	Noun 2	Dif[1)]
English				
JH	*N*	6	6	6
	Mean	147	131	11
	SD	7	4	6
NS1	*N*	6	6	6
	Mean	152	139	31
	SD	36	19	35
NS2	*N*	6	6	6
	Mean	160	122	39
	SD	35	9	31
Japanese				
JH	*N*	7	7	7
	Mean	147	145	7
	SD	9	8	5
Other Japanese texts				
JP1	*N*	15	15	15
	Mean	125	116	10
	SD	17	11	11
JP2	*N*	15	15	15
	Mean	144	136	15
	SD	41	41	13

Note. 1) Dif = $(1/N) \sum | (F_0(\text{Noun 1}) - F_0(\text{Noun 2}) |$.

Table 4

F_0 Values for *camera* and Contexts Where the Word Appeared

S	Function	Context	JH	NS1	NS2
3.	focus	the contraband camera.	129	146	225
5.	new	to a camera club,	138	225	165
6.	new	a better camera.	136	198	137
9.	old	at the camera club once before,	134	141	121
9.	new	such a camera.	136	173	134
12.	new	bought a camera.	132	172	121
15.	old	sold the camera to George Deering,	140	139	144
16.	old	examined the camera,	150	134	158
35.	focus	the contraband camera,	145	125	208
Mean			138	161	157
SD			6	33	37

Table 5

Mean F_0 Values for *he* and *Ben* in the English text

and *kare* and *Ben* in the Japanese text

English

		he	*Ben*	Difference	*t*	*p*
JH	*N*	18	14	32		
	Mean	163	155	8	1.41	*n.s.*
	SD	17	17			
NS1	*N*	13	13	26		
	Mean	143	169	26	1.88	< .08
	SD	20	46			
NS2	*N*	15	13			
	Mean	126	184	58	5.27	< .001
	SD	23	34			

Japanese

		kare	*Ben*	Difference	*t*	*p*
JH	*N*	7	16	23		
	Mean	156	154	2	0.54	*n.s.*
	SD	11	8			

Appendix

The first three paragraphs used on the oral reading task.

Case of the Contraband Camera

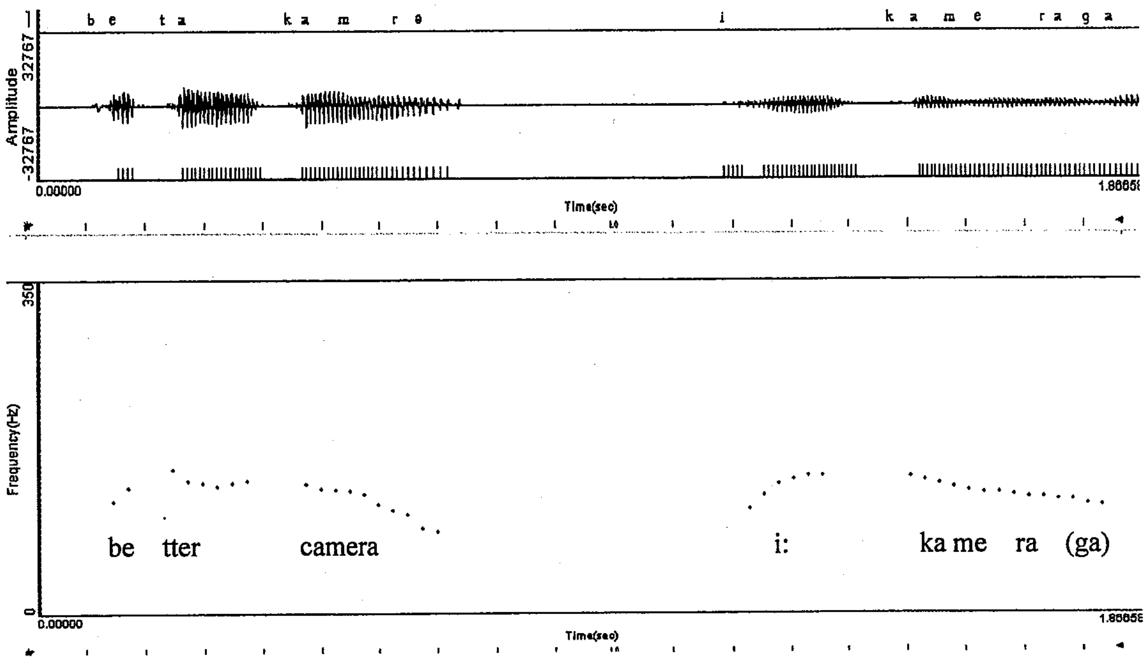
The average hobbyist is a kind of scholar. At least he is always trying to improve his skill and usually his equipment. But sometimes this worthy purpose has unexpected results—such as came out in the CASE OF THE CONTRABAND CAMERA.

Ben Hammer was an amateur photographer who was quite serious about his hobby. He belonged to a camera club, participated in some of their exhibits, and had won a minor award or two for some of salon prints.

Ben felt he'd have made a more impressive showing if he owned a better camera. But the kind he wanted, with the necessary equipment, cost a little over a thousand dollars. Ben, who was a shipping clerk in a wholesale house, didn't have that kind of money.

Figure 1. Pitch patterns of *good camera*: (a) JH's *good camera* and *ii kamera* and (b) NS1 and NS2's *good camera*.

(a)



(b)

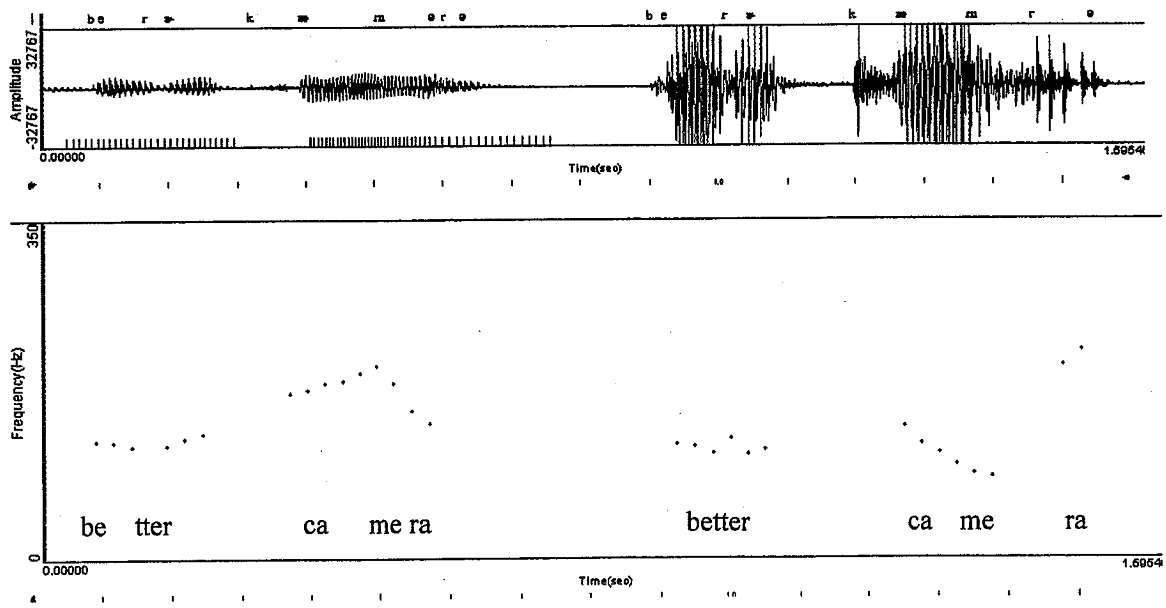
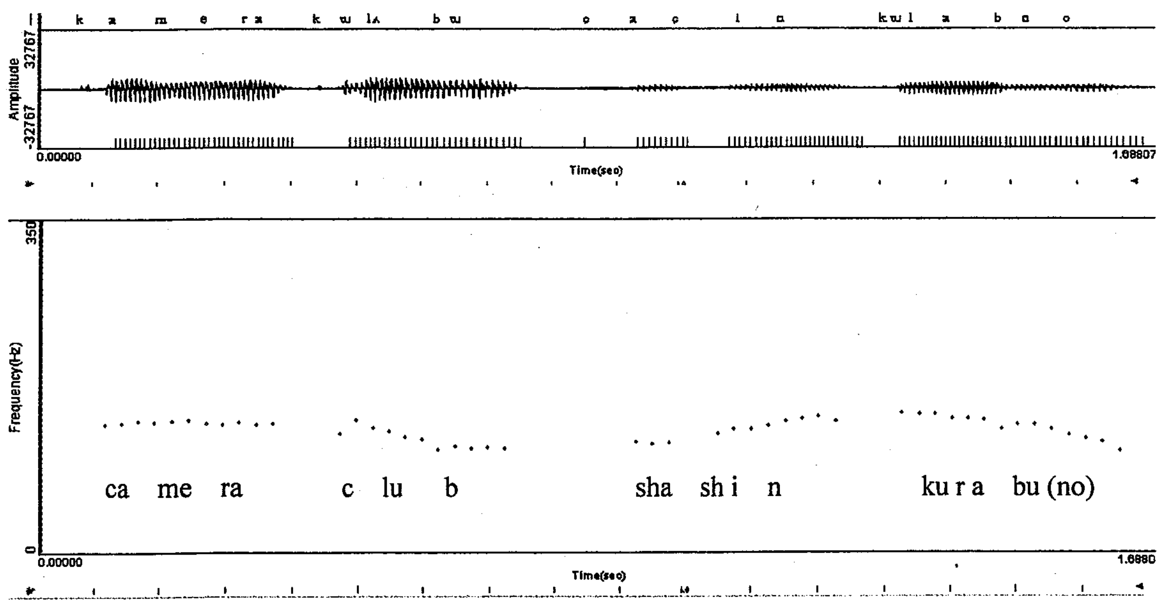
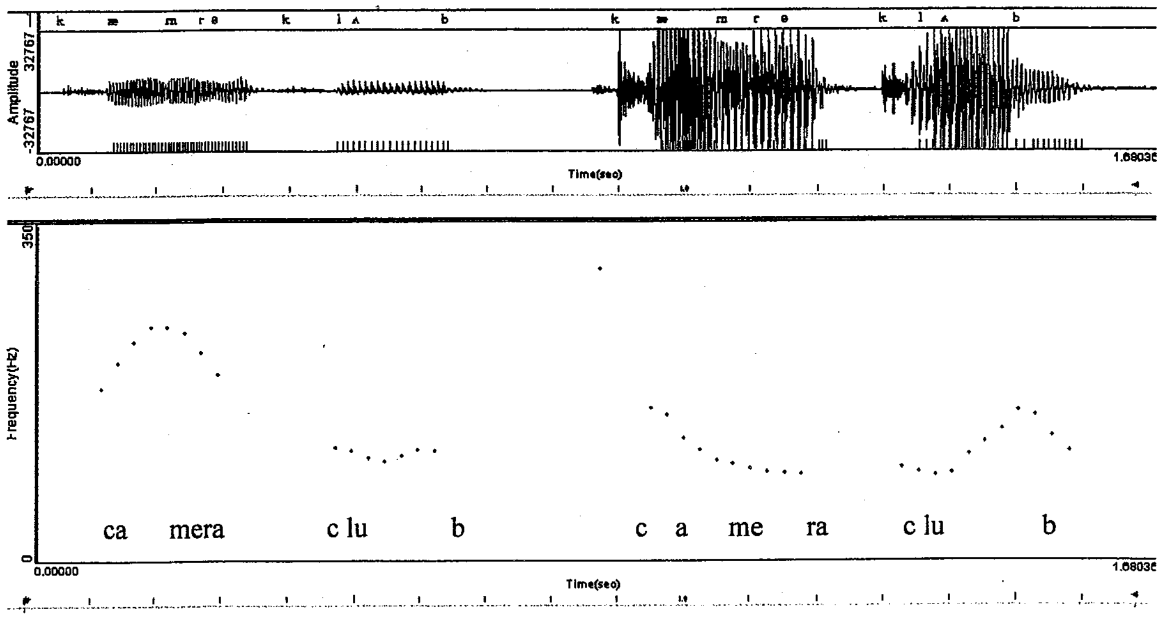


Figure 2. Pitch patterns of *camera club*: (a) JH's *camera club* and *shashin kurabu* and (b) NS1 and NS2's *camera club*.

(a)



(b)



Chapter 2

Mora Duration and Speech Rate in Orally Read Japanese

Abstract

This study measured durations of minimal minor phrases in read speech by a Japanese actor, addressing the issues of mora duration and speech rate. Autocorrelation analysis revealed that the speech rate of a phrase significantly extended to the two immediately following phrases, whereas other results showed that speech rate (or mora duration) greatly varied in a characteristic fashion. Implications of these findings involving the isochrony of Japanese moras are presented.

Introduction

Japanese is traditionally referred to as a mora-timed language where each mora has approximately the same duration. This mora-timing hypothesis seems intuitively obvious, but at the same time, it is equally obvious that a variety of factors such as speech rate, word length, and position in an utterance affect mora duration (see Warner & Arai, 2001, for review). The mora-timing hypothesis would thus remain vacuous unless the effects of various variables on mora duration are explicated. One of the influential variables which affect mora duration is speech rate; or, speech rate can often be synonymous with mora duration. Most previous studies, however, simply assumed that this variable is controlled for simply by instructing subjects to "say the words at a comfortable speaking rate in the carrier sentence," "read the sentences at a normal speaking tempo," "speak the sentences in as normal a manner as possible," and so on. These instructions may not be enough to ensure that speech rate in local environments is kept relatively constant in various experimental conditions. Miller, Grosjean, and Lomanto (1984), for example, showed that there is substantial variation in speaking rate in English within a single utterance of a single speaker in an interview situation. Such intra-speaker variability seems also true about read speech in Japanese although no research has demonstrated that such is indeed the case.

This study measured durations of minimal minor phrases in a read text in Japanese. A minimal minor phrase, defined as a phrase which consists of one content word with or without one or more particles (Poser, 1984), was taken as a minimal unit which reflects a constant speech rate there. This study thus examined how the duration, or speech rate, of minimal minor phrases varies throughout the read text. The effects of word length, which may or may not be related to speech rate, were also determined. Finally, the durations of some of the particle phrases which repeatedly appeared in the read text were examined.

Method

Materials

The subject was a professional actor (M. T.), a native speaker of Tokyo Japanese, and his oral reading on commercially available CD (2002, Shincho-sha, Tokyo) was analyzed. He was 60 years old as of 2002. The text selected was entitled *Lemon*, an essay written by Motojiro Kajii (1901-1932) in 1931, which is commonly used as reading material for high school and college students. The text was 5,738 moras in length and the total oral reading time was about 17 minutes. How the recording was made (e.g., how many rehearsals were made) was not known, but we can assume that it is a model oral reading in Japanese.

The essay describes how the author, depressed by an unknown, evil mass in his mind, bought a lemon at a cozy fruit store, and then left it in Maruzen, a book store which he used to love to visit. Because the events and scenes described there are, in general, low-key, and less emotional, the effects of emotion on speech rate do not seem great.

Measurements

The duration of each minimal minor phrase was measured to the nearest millisecond with reference to the waveform and spectrogram using a Sugi speech analyzer (ANMSW-SSA0101). Where minimal minor phrases consisted of lexical words and one or more particles, the durations of the lexical words and the particles were separately measured, and then combined to obtain the total durations of the minimal minor phrases. This was done to examine possible effects of word class and/or word length on duration.

Results and Discussion

The mean mora duration calculated based on minimal minor phrase durations (i.e., each minimal minor phrase duration divided by the number of moras in it) was 125 msec (N = 1477, SD = 21), showing a coefficient of variation of 17% (SD/Mean × 100). (Note that we hereafter use the mean mora duration as the one calculated this way, unless indicated otherwise.) Although the standard deviation did not appear large enough to suggest that speech rate greatly varies across minimal minor phrases, mora durations ranged from 60 to 259 msec (i.e., from 1,000 to 232 moras per min). The mean duration of the shortest quartile (N= 328) was 102 msec (SD = 7) and that of the longest quartile (N= 328) was 153 msec (SD= 19), the ratio being 1 to 1.5.

Autocorrelations

While a general question was how speech rate varies across minimal minor phrases, a more specific and easier question was to what extent a given mean mora duration (or speech rate) of a minimal minor phrase was associated with those of the preceding and following minimal minor phrases. To answer this latter question, autocorrelations of lag 1 to lag 4 pairs were computed. The results are presented in the first row in Table 1. As is seen from the table, correlations were significant between three adjacent minimal minor phrases, the correlation between first two adjacent phrases being higher than the correlation between the first and the third phrase. This suggested that a given minimal minor phrase 'inherits' its speech rate from the two immediately preceding phrases and extends its speech rate to the two immediately following phrases. These results are surprising in the face of many variables such as word length and final-lengthening (or final-shortening) which disturb the isochrony of moras (e.g., Port, Dalby, & O'Del, 1987; Sato, 1995).

[INSERT TABLE 1 NEAR HERE]

Differences in duration between two adjacent minimal minor phrases, which reflect changes in speech rate, were found interesting. The mean of the absolute values of the differences was 21 msec (N = 1,476, SD = 18). Autocorrelations computed for those differences are presented in the second row in Table 1. The negative correlation here was also unexpected given the positive correlations for the durations (or speech rates) for adjacent minimal minor phrases. This result indicates that speech rate tends to increase and slow down cyclically between adjacent minimal minor phrases. That is, if speech rate becomes faster from a minimal minor phrase (P_i) to the following phrase (P_{i+1}), then it tends to become slower from P_{i+1} to P_{i+2}, and then it tends to become faster from P_{i+2} to P_{i+3}, and so on.

Word length

Inspection of the durations of lexical words and particles within minimal minor phrases suggested that duration was longer for particles than lexical items. This finding allows at least three interpretations: word-length effect, word-class effect, and phrase-final-lengthening effect. Of the three, word length seems to be the strongest. It is well established that, in experimental conditions where speakers carefully articulate test words, the mora duration tends to become somewhat shorter as the word length increases (Port *et al.*, 1987). Thus, regarding the minimal minor phrases with one- or more-mora particles, mean mora durations were determined for one- to six-mora lexical words and those for the co-occurring particles. The results are presented in Table 2.

[INSERT TABLE 2 NEAR HERE]

One remarkable finding was that there was no significant difference in duration between one-mora lexical words and the co-occurring one-mora particles, thereby rejecting the word-class effect hypothesis under the one-mora condition. In contrast, two- or more-mora lexical words were significantly shorter in duration (thus faster in speech rate) than the respectively co-occurring two- or more-mora particles. It is noted that the effect of word length was significant for the two- or more-mora lexical words, whereas that for the two- or more-mora particles was not. Thus, the differences in duration between lexical words and particles seem to be attributable to the effects of final lengthening rather than to those of word class.

To clarify the effect of word length on duration for lexical words, durations of lexical words which constitute minimal minor phrases with no particles followed were determined. The mean was 132 msec (N = 205, SD = 30) for two-mora lexical words, 126 msec (N = 198, SD = 22) for three-mora lexical words, 120 msec (N = 148, SD = 18) for four-mora lexical words, 119 msec (N = 80, SD = 16) for five-mora lexical words, and 119 msec (N = 42, SD = 13) for six-mora lexical words. The effect of word length was significant, $F(4, 668) = 9.35$, $p < .001$.

Given this effect of word length, the possibility arises that the findings concerning the significant correlations between adjacent minimal minor phrases presented in Table 1 might be an artifact because word length is highly correlated with minimal minor phrase length. Lengths of adjacent minimal minor phrases may compensate for each other such that a longer minimal minor phrase tends to be followed by a shorter minimal minor phase and vice versa. To examine this potentially confounding factor, autocorrelations were computed for lengths of minimal minor phrase measured by number of moras. The results showed that the correlations were .05, -.04, .04, and .03 for the lag 1 pairs through lag 4 pairs, none reaching significance. Thus the findings reported in Table 1 were not an artifact.

Durations of one-mora particles

The frequency of function words such as case particles is generally high, and such particles can be used to show how speech rate varies within the same speaker. If speech rate is kept relatively constant, the variability of the durations of such particles should also be kept small. Table 3 summarizes results for five major case particles.

[INSERT TABLE 3 NEAR HERE]

It is apparent from Table 3 that speech rate for the one-mora particles varied substantially. The mean durations ranged from 129 to 160 msec with SDs ranging from 28 to 42 msec, yielding a range of coefficients of variation from 21 to 26%. These values for coefficients of variation are nearly comparable to that reported by Miller et al. (1984) for English speakers' utterances in interviews (i.e., 27%). The ratios of the maximum to minimum durations were even more striking, ranging from 2.5 to 4.3.

Speech rate varied greatly for even the same minimal minor phrases. The phrase /wataʃi wa/ (the first personal pronoun /wataʃi/ "I" followed by the case particle /wa/), for example, appeared 39 times. The mean mora duration was 123 msec with *SD* = 16, *Max* = 162, and *Min* = 95. The ratio of the maximum to minimum duration was 1 to 1.7.

Some issues involving the isochrony of Japanese moras

The findings of this study raise many issues. Only two of them are briefly mentioned here. First, the substantial local variation in speech rate found in this study suggests that unless such local variation in speech rate is controlled for in various experimental conditions, the findings of mora durations in Japanese would be unreliable. It would be necessary to measure many tokens for each test word whether the word is embedded in a carrier sentence (e.g., Beckman, 1982) or in a text (e.g., Campbell, 1992, 1999).

Second and more important, a crucial question arises as to what the underlying mechanisms which control speech rate are. Miller et al. (1984), for example, suggested that for natural conversation in English, the lexical access difficulties, syntactic construction delays, and semantic planning problems can be candidate factors. But as far as oral reading is concerned, none of them would exert a great influence on speech rate because all are already presented out in the text. Rather, aside from linguistic factors, the cyclic variation in speech rate shown in Table 1 seems to suggest the existence of a central pattern generator which modulates speech rate in a short range. It is unknown if such a central pattern generator is somehow associated with those postulated for locomotion, respiration, and mastication in vertebrates (cf. Cohen, Rossignol, & Grillner, 1988; Hausdorff et al., 1995). Intra- and cross-linguistic research on local variation in speech rate, which is needed to examine the generality of the present results, may also shed light on this general issue.

References

Campbell, N. (1992). Segmental elasticity and timing in Japanese. In Y. Tokura, E. Vatikiotis-Bateson, and Y. Sagisaka (Eds.). *Speech, perception, production, and linguistic structure*, pp. 403-418. Tokyo: Ohm-sha.

Campbell, N. (1999). A study of Japanese speech timing from the syllable perspective. *Journal of the Phonetic Society of Japan*, 3, 29-39.

Cohen, A. H., Rossignol, S., and Grillner, S. (1988). *Neural control of rhythmic movements in vertebrates.* New York: John Wiley & Sons.

Hausdorff, J. M., Peng, C. –K., Landin, Z. et al. (1995). Is walking a random walk? Evidence for long-range correlations in stride interval of human gait. *Journal of Applied Physiology*, 78, 349-358.

Miller, J. L., Grosjean, F., and Lomanto, C. (1984). Articulatory rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, 41, 215-225.

Port, R. F., Dalby, J., O'Del, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustic Society of America*, 81, 1574-1585.

Poser, W. J. (1984). *The phonetics and phonology of tone and intonation in Japanese.* Doctoral dissertation, MIT.

Sato, Y. (1995). The mora timing in Japanese: A positive liner correlation between the syllable count and word duration. [In Japanese with English summary]. *The Bulletin of the Phonetic Society of Japan*, 209, 40-53.

Warner, N. and Arai, T. (2001). Japanese mora-timing: A review. *Phonetica*, 58, 1-25.

Table 1

Autocorrelations for the Mean Mora Durations of Minimal Minor Phrases (First Row, $df = 1,473$) and those for the Differences in Duration between Adjacent Minimal Minor Phrases (Second Row, $df = 1,472$)

	Lag 1	Lag 2	Lag 3	Lag 4
Mora Durations	.13**	.08*	.03	-.00
Differences	-.47**	.00	-.01	-.04

*$p < .01$, **$p < .001$.

Table 2

Mean Mora Durations and Standard Deviations

for One- to Six-mora Lexical Words and the Co-occurring Particles

		Lexical Words		Particles			
Word length	*N*	*Mean*	*SD*	*Mean*	*SD*	*t*	*p*
One-mora	13	145	36	131	31	0.19	*n.s.*
Two-mora	232	124	22	137	33	5.33	< .001
Three-mora	236	116	17	144	40	10.46	< .001
Four-mora	153	112	14	144	33	11.53	< .001
Five-mora	35	110	11	144	36	5.42	< .001
Six-mora	23	108	11	138	38	3.28	< .01

Table 3

Speech Rate Variation Measures for Five Major Case Particles: *N* = number of occurrences, *Mean* = mean duration (msec), *SD* = standard deviation; *Max* = maximum duration, and *Min* = minimum.

Case particle	*N*	*Mean*	*SD*	*Max*	*Min*
Subjective /wa/	85	160	42	270	96
Subjective /ga/	81	138	36	279	78
Objective /o/	81	138	35	239	55
Objective /ni/	95	129	30	218	80
Possessive /no/*	131	134	28	216	88

*The typical function of /no/ is possessive, but other functions are also included in this category.

Chapter 3

Normal Japanese Speakers' Nasalance Values: The Effects of Speaker Subgroup and Text Passage

Objective: To investigate sources of apparent inconsistent findings concerning cross- and intra-linguistically compared nasalance values for normal speakers.

Subjects: Normal Japanese speakers with basic knowledge of English as a second language.

Materials: English and Japanese passages. In English, the high-pressure Zoo Passage and the Nasal Sentences. In Japanese, the high-pressure Woodpecker and low-pressure Blue House Passages, and the Sparrow Passage. In addition, the low-pressure Top Passage was given to half of the subjects.

Results: The mean nasalance scores for the six passages were 12.5%, 52.5%, 10.8%, 18.1%, 50.8%, and 13.6%, respectively. Inspection of results suggested the existence of two types of low-pressure oral passage and three subtypes of speakers.

Conclusion: The apparent inconsistencies in previous studies are largely attributable to the effects of speaker subgroup and text passage.

KEY WORDS: *nasalance, Japanese, intraoral pressure, subgroups of normal speakers*

Introduction

Nasalance values, the ratios of nasal acoustic energy to the sum of nasal plus oral acoustic energy, are useful in assessing persons at risk for velopharyngeal insufficiency. Those values measured by the Nasometer (Kay Elemetrics) for oral speech, however, appear to vary cross- and intra-linguistically even among normal speakers. Mayo et al. (1996) used the Zoo Passage (a standard oral passage in English) and found a mean nasalance score of 17.0% (SD = 5.2) for African-Americans and a mean of 18.2% (SD = 4.2) for Caucasian-Americans. For oral speech in other languages, Nichols (1999) reported 15.5% (SD = 15.5) for Spanish speakers living in Mexico City and 18.7% (SD = 7.4) for those living in Cuernavaca (a city to the south of Mexico City), whereas Anderson (1996) obtained 22% (SD = 8.7) for Spanish-speaking women; Hirschberg et al. (2006) found 13.4% for Hungarian speakers, Müller et al. (2000) found 13.0% for German speakers, and Whitehill (2001) found 16.8% (SD = 6.0) for Cantonese-speaking women. Toward the lower end of the nasalance scale, van Lierde et al. (2001) obtained 10.9% (SD = 4.2) for young Flemish adults, and Tachimura et al. (2000), 9.1% (SD = 3.9) for Japanese speakers. We note that the mean for the Japanese group in Tachimura et al. (2000) was significantly smaller than that for the Flemish group in Van Lierde et al. (2001), $t(156) = 2.69$, $p < .01$.

A question naturally arises as to where this apparently wide range of nasalance values comes from. In particular, we ask why Japanese speakers have such small nasalance values as compared with other language speakers. There can be several plausible explanations[1)]. Tachimura et al. (2000), for example, suggested two: morphological characteristics of the Mongolian people and Japanese syllable structure. The first explanation is immediately refuted given Whitehill's (2001) finding that Cantonese speakers had a mean nasalance value of 16.8%. As for the second possibility, Tachimura et al. (2000) speculated that because Japanese sentences largely consist of vowel (V) and consonant-vowel (CV) syllables, more vowels included in the passage would increase oral energy and thus decrease nasalance scores. This speculation, however, is inconsistent with what Tachimura and associates (Hirata et al. 2002) found more recently. That is, Hirata et al. (2002) found high mean nasalance scores for V and CV syllables pronounced in isolation by normal Japanese speakers, e.g., 39.6% for /i/, 32% for /bi/, and 25.8% for /e/. Also, surprisingly, they obtained a mean of 6.8% for the low-pressure Top Passage, i.e., "Ue wo oou. Yooi wa ooi" (Cover the top. Preparations are many.), and a mean of 9.8 for the high-pressure Knock-knock Passage, "Kotsu kotsu tsutsuku. Kitsutsuki tsutsuku" (Knock-knock sounds. The woodpecker pecks.).

In this regard, equally surprising results were reported by Ogata et al. (2003), who measured nasalance scores for Japanese children with and without velopharyngeal insufficiency. Ogata et al. (2003) used Hirata et al.'s (2002) Top and Knock-knock Passages, and the mean nasalance scores for normal children (N = 20) were 8.4% (SD = 4.2) for the former passage and 16.2% (SD = 6.6) for the latter. These findings are difficult to interpret. The mean of 16.2% for the high-pressure Knock-knock Passage is far greater than the mean of 9.1% reported by Tachimura et al. (2000) for the high-pressure Woodpecker Passage (see the Appendix). Basically we would expect nasalance to be *greater* for a low-pressure passage than for a high-pressure passage. The reason is because stops and affricates appearing in a high-pressure passage normally require complete velic closure and thus greater intraoral air pressure (e.g., Schourup, 1972). Hirata et al.'s (2002) finding about normal adult speakers was also against our expectation; namely, it showed a mean of 6.8% for the low-pressure passage and a mean of 9.8% for the high-pressure Knock-knock Passage, the difference being significant. Note that neither Ogata et al. (2003) nor Hirata et al. (2002) discussed this issue, probably because their main concern was nasalance scores for patients with velopharyngeal insufficiency. On the other hand, for English-speaking patients with (marginal) velopharyngeal insufficiency, Karnell (1995), Karnell et al. (2001), and Watterson et al. (1998) found complicated patterns of high-pressure and low-pressure nasalance scores. For example, Karnell (1995) suggested, albeit considered preliminary, that for some patients, velopharyngeal closure for vowels and semivowels is adequate while for others, that for pressure consonants is relatively adequate. To date, however, the present author is not aware of relevant normative data concerning nasalance values for high- and low-pressure passages among normal speakers of English or other languages.

This study was conducted to search for sources of (1) differences in nasalance scores between Tachimura et al.'s (2000) Japanese speakers and non-Japanese speakers in previous studies, and (2) differences in findings among Tachimura et al. (2000), Ogata et al. (2003), and Hirata et al. (2002). Regarding the first issue, normal Japanese speakers with some knowledge of English were asked to read both the Zoo Passage and the Woodpecker Passage, and their nasalance values for these passages were compared with those of Tachimura et al.'s subjects and English speakers. 'Bilingual' speakers may provide us with a useful paradigm in which to investigate the cross-linguistic effects. As to the second issue, the same subjects were asked to read another oral passage, i.e., the low-pressure Blue House Passage, which was composed of only vowels and semivowels. The question was if we could replicate the pattern of findings reported by Ogata et al. (2003) and Hirata (2002) when we use a new low-pressure passage. In addition, two nasal passages were used to collect more comprehensive data.

METHOD

Subjects

Forty-seven Hiroshima University students (25 males and 22 females) served as subjects. The subjects ranged in age from 19 to 24 years with a mean of 19.8 years (SD = 2.0). All were native Japanese speakers with normal speech and had been learning English as a second language for more than six years largely in classroom settings. Many of them were not fluent speakers of English, but all were able to read easy English such as the Zoo Passage. Most of them were from various places in Western Japan.

Reading Stimuli

Two English passages, the Zoo and Nasal Sentences Passages, and three Japanese passages, the Woodpecker, Sparrow, and Blue House Passages, were used as reading materials (see the Appendix). For the Zoo Passage, the first six sentences of the standard passage were employed (cf. Karnell, 1995; Watterson et al. 1999; Watterson et al., 2001). There were no nasal consonants in the Zoo, Woodpecker, and Blue House Passages, whereas the Nasal sentences and Sparrow Passages were loaded with nasal consonants. The Sparrow and Blue House Passages were constructed by the present author, taking the comparable passages of previous studies into consideration[2)]. The Japanese passages presented to subjects were written in standard form using *kanji* (Chinese characters), *hiragana*, and *katakana*.

Equipment

The nasalance scores presented in the introduction were obtained with the first version of Nasometer (Kay Elemetrics). In this study, a new version, the Model 6400-II Nasometer (KayPENTAX), was employed. The Nasometer II Model 6400 manual reads: "The normative data obtained on Nasometer II varies somewhat from data reported on the original Nasometer" (p. 2). The manual (p. 59) also reports that the mean score for the Zoo Passage was 11.3% (SD = 5.6) for 40 normal speakers of English. Although this mean is lower than those reported by Seaver et al. (1991) and Mayo et al. (1996), it is not readily apparent to what extent this difference is attributed to the difference between the old Nasometer and the Nasometer II. Watterson et al. (2005) compared the old and new machines and found that "[the] overall variability was not so great as to cause concern about applying normative data obtained previously with the old Nasometer to nasalance scores obtained with the Nasometer II" (p. 579) although they also stated that care should be exercised when comparing nasalance scores between the two machines.

Nasometer calibration was checked frequently enough (though not every time prior to testing). The headgear was not adjusted or replaced between readings.

Procedure

Each subject was seated in a comfortable chair in a quiet (but not sound-treated) room. The headgear was quickly and properly adjusted following the instructions provided by the KayPENTAX Nasometer manual. The subject first read a test passage silently and, in the case of English, was asked if there were any words which he/she did not know. Many subjects were not very sure about the pronunciations of "fluffy," "Bounding," and

"Maine," and thus the examiner modeled the words. Each subject then orally read the Zoo Passage first, followed by the Nasal Sentences, Woodpecker, Sparrow, and Blue House Passages in that order. After reading each passage, the subject had a brief break, during which the examiner loaded the next passage on the computer display.

RESULTS

The main results of this study and the comparable findings of previous research are summarized in Table 1. For the oral passages in this study, the effect of passage was significant, $F(2, 92) = 40.62$, $p < .001$. The mean for the Blue House Passage (18.1%) was significantly greater than the mean for the Zoo Passage (12.5%), $t(46) = 6.84$, $p < .001$, which in turn was significantly greater than that for the Woodpecker Passage (10.8%), $t(46) = 2.76$, $p < .01$.

[TABLE 1 GOES NEAR HERE]

The mean for the high-pressure Woodpecker Passage (10.8%) in this study tended to be greater than the means for the same passage (9.1%) in Tachimura et al.'s (2000) study, $t(145) = 1.98$, $p = .05$, but was not significantly different from the mean for the high-pressure Knock-knock Passage (9.8%) in Hirata et al. (2002), $t(56) = 0.64$. These three means, however, were significantly smaller than the mean for the Knock-knock Passage reported by Ogata et al. (2003), e.g., $t(65) = 3.23$, $p < .01$, for the present study vs. Ogata et al (2003). On the other hand, the mean for the Zoo Passage in the current study was not significantly different from that reported by the Nasometer II Model 6400 manual for 40 normal American English speakers (p. 59), $t(85) = 0.36$, n.s.

The pattern of the present result that the difference in mean nasalance values between the low-pressure Blue House Passage (18.1%) and the high-pressure Woodpecker Passage (10.8%) was highly significant, $t(46) = 6.95$, $p < .001$, was in complete conflict with those of Ogata et al. (2003) and Hirata et al. (2002), i.e., 8.4% (the low-pressure Top Passage) vs. 16.2% (the high-pressure Knock-knock Passage), and 6.8% (the low-pressure Top Passage) vs. 9.8% (the high-pressure Knock-knock Passage), respectively. In the current study, only four subjects out of 47 (8.5%) exhibited lower nasalance scores for the low-pressure Blue House Passage ($M = 10.5\%$) than for the high-pressure Woodpecker Passage ($M = 12.3\%$).

The question at issue was what the conflicting findings between the current study, Ogata et al. (2003), and Hirata et al. (2002) are attributed to. This question soon became apparent during data collection, and after testing 20 subjects, inspection of the results for the Blue House Passage revealed that nasalance values were generally lower for the last phrase of the passage (i.e., /uee uee/ (up and up) than elsewhere. This suggested that the low-pressure Blue House Passage might qualitatively be different from the low-pressure Top Passage used by Ogata et al. (2003) and Hirata et al. (2002). Thus the remaining 27 subjects were additionally given the Top Passage. The result was that the mean was 13.6% ($SD = 10.2$), which was still greater than those of Ogata et al. and Hirata et al. The scores of these subjects, however, appeared to exhibit a kind of trimodal distribution, and thus the subjects were divided into three subgroups: the high-scorer subgroup ($N = 9$, $M = 26.8\%$, $SD = 6.1$), the medium-scorer subgroup ($N = 3$, $M = 11.3\%$, $SD = 1.2$), and the low-scorer subgroup ($N = 15$, $M = 6.2\%$, $SD = 1.5$). This last subgroup's mean was very close to 6.8% found by Hirata et al. (2002), $t(22) < 1$, while it was smaller than 8.4% obtained by Ogata et al. (2003). This subgroup was similar to the subjects from Hirata et al. and Ogata et al. in that the mean for the high-pressure Woodpecker Passage was 8.0% (SD = 2.7), which was significantly higher than the mean for the Top Passage, 6.2% (SD = 1.5), $t(14) = 3.83$, $p < .01$.

Finally, regarding the nasal passages, the mean (50.8%) for the Japanese Sparrow Passage was significantly lower than the mean (52.5%) for the English Nasal Sentences Passage, $t(46) = 2.50$, $p < .05$. More interest, the mean for the Nasal Passage was significantly lower than the mean (59.6%, SD = 8.0) for the same passage for 40 native English speakers reported by the Nasometer II Model 6400 manual (p. 59), $t(85) = 4.06$, $p < .001$.

DISCUSSION

The major findings were (1) that for the Japanese high-pressure Woodpecker Passage in Japanese, the present mean nasalance value tended to be greater than that found by Tachimura et al. (2000), (2) that the present mean nasalance value for the standard English low-pressure Zoo Passage was significantly greater than the mean for the Woodpecker Passage, (3) that this mean for the Zoo Passage, however, was not significantly different from normal English speakers' mean for the same passage, (4) that, of the two Japanese oral passages, the mean was the greater for the low-pressure Blue House Passage than for the high-pressure Woodpecker Passage, a pattern which was the opposite of the findings of Ogata et al. (2003) and Hirata et al. (2002), and (5) that, given the low-pressure Top Passage used by those researchers, however, some subjects exhibited the same pattern as the pattern reported by those researchers. We will briefly discuss these findings below.

The first finding concerning a small but significant difference may be due to the different versions of the Nasometer used in this study and Tachimura et al. (2000), and/or to a possible difference in dialect. While the first possibility remains unclear (see Watterson et al., 2005), the second possibility should be ascertained in future research by collecting more data about speakers from different areas in Western Japan.

The second finding that, overall, a high-pressure passage was associated with a low nasalance score is what we expect (e.g., Schourup, 1972), but this finding may not be characterized as a norm not only because previous studies reported a different pattern but also because this study suggested the existence of sub-groups who exhibit different patterns (see below).

The third finding that the mean nasalance values for the Zoo Passage were not significantly different between the present Japanese speakers and the normative English speakers does not support Tachimura et al.'s (2000) claim that Japanese speakers uniquely demonstrate lower nasalance scores for oral speech as compared with other language speakers. Tachimura et al.'s speculation that "the normative nasalance data previously reported for the Zoo passage may not be valid for Japanese speakers" (p. 465) is thus unacceptable; instead, we conclude that the Zoo Passage is valid not only for English speakers but also for Japanese speakers and probably for other speakers. We also conclude that it is not the Japanese unique phonological system (if at all) but rather phonological properties of different oral passages used in studies of speakers of different languages that elicit different nasalance values.

As for the fourth finding of this study, a baffling question remained. We wondered why Ogata et al. (2003) and Hirata et al. (2002) found smaller nasalance values for the low-pressure Top Passage than for the high-pressure Knock-knock Passage. As stated in the introduction, these findings are not consistent with our natural expectation that a high-pressure passage should be associated with a lower nasalance value because of more adequate velic closure and greater intraoral pressure required in such a passage.

At least a partial solution to this question, however, .The fifth finding regarding the different nasalance values across subjects and passages is of importance in that it suggests While a low-pressure oral passage may be further classified into two subtypes, i.e., one which elicits higher nasalance values and the other, lower nasalance values, three subgroups of speakers who differentially responded to the Top Passage were identified: the high-, medium-, and low-scorer subgroups. The low-scorer subgroup (N = 15) resembled the subjects of Hirata et al. (2002) and Ogata et al. (2003) in that all these subjects showed lower nasalance values for the Top Passage and higher values for the high-pressure passages. The other subgroups produced higher values not only for the Top Passages but also for the Blue House Passage; 26.8% (SD = 6.1) and 29.7% (SD = 8.3) for the high-scorer subgroup (N = 9), and 11.5% (SD = 1.0) and 18.8% (SD = 1.5) the medium-scorer subgroup, respectively. It thus suggested that it is the distribution of the subgroups that can affect overall means of nasalance values.

The existence of subgroups raises a new question: How are such subgroups formed. Dialect and/or idiolect may be an answer. Normal speakers may somehow differentially acquire open nasalized vowels to different degrees. Concerning English speakers, for example, Catford (2001) observes that "[s]ome people nasalize very open vowels, like [ɑ] all the time, saying [pã] [spã] [kã(r)], etc. for pa spa car" (p. 75). The same may apply to Japanese speakers. However, which vowels tend to be more nasalized should be investigated in further research.

Finally, some clinical implications may be suggested on the basis of these findings and interpretations. First, because nasalance values vary across oral passages, it is recommended to use two or more different types of oral passages. Second, the characteristics of subgroups of speakers in terms of the interaction between nasality and intraoral pressure should be elaborated in future research. Specifically, the question should be addressed as to whether there are subgroups in Japanese patients with VPI analogous to those suggested in this study (cf. Karnell, 1995; Karnell et al., 2001). Third, individual differences in normal speakers can be larger than generally expected. There were several subjects who had high nasalance values for oral passages; for example, three subjects' nasalance values for the Blue House Passage were 35% or higher. The question may remain how we can identify such speakers as normal. Finally, results from the nasal passages may be informative. The finding that the mean for the Nasal Passage in this study was significantly lower than the mean for the same passage for native English speakers reported by the Nasometer II Model 6400 manual (p. 59) can be taken to claim that Japanese speakers' nasals are generally less nasal than English speakers'. This point is consistent with the possibility that some amount of air is likely to flow out of the mouth when Japanese speakers produce /n/ of "none," for example. (Because of this, Japanese speakers have no difficulty saying words such as "none" with their nostrils pinched firmly.) Given these possibilities, it may be more effective to use not only different types of oral passages but also some nasal passages for clinical purposes.

Notes

1) A reviewer of this journal has pointed out that an obvious reason is that the equipment may not be reliable. For example, Zajac et al. (1996) describe

variability in the nasalance scores between different Nasometer microphones. Measurement reliability of the Nasometer can also be a problem as Bressmann (2005) identified nasalance test/retest values raging from 4 to 6%. These fundamental problems, however, are beyond of this study, and we have to bear them in mind when we compare findings between different studies. Brief mention of the possible effect of the versions of the Nasometer is made in the method section.

2) A reviewer of this journal asks if the nasal Sparrow Passage is comparable to the Nasal Sentences. The percentage of nasal phonemes in the Sparrow Passage is 27%, which is lower than the 35% in the Nasal Sentences. But because the Sparrow Passage includes a moraic nasal and devoices vowels, it is not readily apparent that the two nasal passages differ in terms of nasal density. Rather, mean nasalance values for these passages would show us how they differ.

References

Anderson RT. Nasometric values for normal Spanish-speaking females: A preliminary report. *Cleft Palate Craniofac J* 1996;33:333-336.

Bressmann T. Comparison of nasalance scores obtained with the Nasometer, the NsalView, and the OroNasal System. *Cleft Palate Craniofac J* 2005;42:423-433.

Catford JC. A practical introduction to phonetics, second edition. Oxford: Oxford, 2001.

Dalston RM, Neiman GS, Gonzalez-Landa G. Nasometer sensitivity and specificity: A cross-dialect and cross-cultural study. *Cleft Palate Craniofac J* 1993;36:57-63.

van Doorn & Purcell. Nasalance levels in the speech of normal Australian children. *Cleft Palate Craniofac J* 1998;35:287-292.

Hirata S, Wada T, Tachimura, T, Hara, H, Nohara, K, Tatoh K. Nasalance of speech sample in Kansai (the mid-western part of Japan) dialect speakers and its availability as an evaluation tool for identifying patients with velopharyngeal incompetence following cleft palate surgery. (In Japanese with English abstract). *J Jpn Cleft Palate Assoc* 2003; 27:14-23.

Hirschberg J, Bok S, Juhasz M, Trenovszki Z, Votisky P, Hirschberg A. Adaptation of nasometry to Hungarian language and experiences with its clinical application. *Int J Pediatr Otorhinolaryngol* 2006;70:785-798.

Karnell MP. Nasometric discrimination of hypernasality and turbulent nasal airflow. *Cleft Palate Craniofac J* 1995;32:145-148.

Karnell MP, Schultz K, Candy J. Investigations of a pressure-sensitive theory of marginal velopharyngeal inadequacy. *Cleft Palate Craniofac J*

2001;38:346-357.

van Lierde KM, Wuyts FL, de Bodt M, van Cauwenberge P. Nasometric values for normal nasal resonance in speech of young Flemish adults. *Cleft Palate Craniofac J* 2001;38:112-148.

Mayo R, Floyd LA, Warren DW, Dalston RM, Mayo CM. Nasalance and nasal area valus: Cross-racial study. *Cleft Palate Craniofac J* 1996;33:143-149.

Müller R, Beleites T, Hloucal U, Kühn M. Objektive Meassung der normalen Nasalanz im sachsischen Sprachraum. *HNO* 2000;48:937-942.

Nichols AC. Nasalance statistics for two Mexican populations. *Cleft Palate Craniofac J* 1999;36:57-63.

Ogata Y, Nakamura N, Kubota Y, Sasakuri M, Kikuta R, Shirasuna K, Oish M. Nasometer test for analysis on velopharyngeal function of patients with cleft plate: Objective criteria for velopharyngeal function. (In Japanese with English abstract). *J Jpn Cleft Palate Assoc* 2003; 28:9-19.

Schourup L. Characteristics of nasal vowels. *Papers in Linguistics* 1972;5: 530-548.

Seaver EJ, Dalston RM, Leeper HA, Adams LE. A study of nasometric values for normal nasal resonance. *J Speech Hear Res* 1991;34:715-721.

Tachimura T, Mori C, Hirata S, Wada T. Nasalance score variation in normal adult Japanese speakers of Mid-West Japanese dialect. *Cleft Palate Craniofac J* 2000;37:463-467.

Watterson T, Lewis KE, Brancamp T. Comparison of nasalance scores obtained with the Nasometer 6200 and the Nasometer II 6400. *Cleft Palate Craniofac J* 2005;42:574-579.

Watterson T, Lewis KE, Deutsch C. Nasalance and nasality in low pressure and high pressure speech. *Cleft Palate Craniofac J* 1998;35:293-298.

Watterson T, Lewis KE, Foley-Homan N. Effect of stimulus length on nasalance scores. *Cleft Palate Craniofac J* 1999;36:243247.

Whitehill TL. Nasalance measures in Cantonese-speaking women. *Cleft Palate Craniofac J* 2001;38:119-125.

Zajac, DJ, Lutz, R, Mayo, R. Microphone sensitivity as a source of variation in nasalance scores. *J Speech Hear Res* 1996;39:1228-1231.

Table 1

Mean and SD Nasalance Values (%) for the High- and Low-Pressure Oral and Nasal Passages in This and Previous Research

	Mean	SD
High-Pressure Oral Passage		
In English		
This study: Zoo	12.5	6.4
Nasometer II manual: Zoo	11.3	5.6
In Japanese		
This study: Woodpecker	10.8	5.3
Tachimura et al. (2000): Woodpecker	9.1	3.9
Hirata et al. (2002): Knock-knock	9.8	4.3
Ogata et al. (2003): Knock-knock	16.2	6.6
Low-Pressure Oral Passage		
In Japanese		
This Study: Blue House	18.1	8.8
This study (N = 27): Top	13.6	10.2
Hirata et al. (2002): Top	6.8	3.6
Ogata et al. (2003): Top	8.4	4.2
Nasal Passage		
In English		
This study: Nasal Sentences	52.5	8.3
Nasometer II manual: Nasal Sentences	59.6	8.0
In Japanese		
This study: Sparrow	50.8	8.0

APPENDIX

Oral passage

(1) Zoo Passage:

Look at this book with us. It's a story about a zoo. That is where bears go. Today it's very cold out of doors. But we see a cloud overhead. That's a pretty white fluffy shape.

(2) Woodpecker Passage (Tachimura et al. 2000):

Kitsutuki ga ki wo tsutsuku. Suku suku sodatsu. Te wo tataku. Te ga todoku. (A woodpecker pecks at a tree. Children grow up quickly. We clap hands. A hand reaches it.)

(3) Blue House Passage

Aa aoi ie. Awai ao wa yowai. Ai wa ii. Ayu wa ue e ue e. (Oh, blue houses. Light blue is weak. Love is nice. Ayu fish swim up and up.)

Nasal passage

(1) Nasal Sentences:

Mama made some lemon jam. Ten men came in when Jane rang. Dan's gang changed my mind. Ben can't plan on a lengthy rain. Amanda came from Bounding, Maine.

(2) Sparrow Passage:

Mura no suzume ga minami wo muite naite imasu. Mawari wo mimawasite mo nani mo imasen. (The sparrow in a village is cheeping looking to the south. Looking around the place, we don't see anything.)

Chapter 4

A simple and effective treatment for stuttering: Speech practice without audience

Co-authored with Takanobu Homma

Abstract

On the assumption that stuttering is essentially acquired behavior, it has been concluded that speech-related anticipatory anxiety as a major cause of stuttering accounts for virtually all apparently-different aspects of stuttering on the behavioral level. Stutterers' linguistic competence is unimpaired, although their speech production is characterized as "disfluent." Yet, such disfluency is dramatically reduced when such people speak in anxiety-free no-audience conditions (Bloodstein, 1995). Furthermore, our pilot study of oral reading in Japanese indicates that a stutterer can easily replace stuttering events with a common interjection, "eh," and make oral reading sound natural and fluent. Given these facts, we propose the Overlearning Fluency when Alone (OFA) treatment, consisting of two distinct but overlapping steps: (1) Overlearning of fluency in a no-audience condition, and (2) Use of an interjection, "eh," as a starter when a stuttering event is anticipated. It remains to be demonstrated that this is a truly simple and effective treatment for "one of mankind's most baffling afflictions."

Introduction

Despite the numerous studies over the years, stuttering, defined as an involuntary and intermittent disruption in the fluency of speech production, remains a mystery [1, 2]. The following characteristic aspects of stuttering are well known: Most stuttering occurs in word- or sentence-initial position; in chorus reading, stuttering is dramatically reduced; it is also greatly reduced when the stutterer is alone; it is also greatly reduced in repeated oral reading; isolated words are stuttered less often than words in sentences; there are about four times more stutterers among males than females; ~80% of developmental stuttering is spontaneously recovered; feared sounds and words are easier to produce when first saying “ah” or using some other ‘starter’; and in singing, almost no stuttering occurs.

The mystery of stuttering lies in the cause(s) of such a varied phenomenon. In this study, we first attempt to reformulate a basic hypothesis with which to account for those phenomena on the behavioral level. We view speech-related anticipatory anxiety as a major inhibitor of fluent speech, and a starter as a good inhibitor of stuttering. We then propose what we call the Overlearning Fluency while Alone (OFA) treatment on the basis of this hypothesis.

Accounts of stuttering phenomena

The probable causes of stuttering at the behavioral level were extensively discussed in the early decades of the twentieth century [1, 3], and in line with Johnson and others [3], we assume that stuttering is essentially acquired behavior. A role model from whom the to-be stutterer learns to stutter is either a person around him (a parent, grandparent, sibling, friend, etc.) or himself. We also assume that stutterers’ linguistic competence is not impaired. The simple fact that stutterers speak fluently when they are alone is sufficient enough to verify these assumptions. Thus, like Johnson and colleagues, we propose that a chief underlying cause of stuttering is social speech-related anticipatory anxiety. In other words, to answer the question of why stuttering is greatly reduced when alone, we simply refer to the fact that the stutterer's true linguistic competence best manifests itself in such an anxiety-free condition.

Although the striking effect of no audience was found in the late 1930s [3, 4, 5], this finding seems to have got buried in the literature with sporadic research focusing on audience size with one or more listeners [6, 7]. We have rediscovered the impressive effect of no audience when the second author, a Japanese stuttering researcher who stutters in a mild-moderate degree, served as a subject in an oral reading experiment in an audience condition where the first author was the only audience and in a no-audience condition. He read aloud several short high-school-level passages in these conditions. The difference between the two conditions was striking. His oral readings in the no-audience condition were tension-free, natural sounding, and fluent except for occasional repetitions. The mean reading rate for one passage (No. syllables = 532), for example, was 227 syllables per minute (SPM) in the no-audience condition, and 115 SPM in the audience condition, the former being about two times faster than the latter; equally dramatic was the mean frequency of stuttering per 100 syllables, i.e., 2.8 in the no-audience condition and 10.3 in the audience condition, the ratio being 1 to ~3.7. Regarding fundamental frequency (F_0), which is considered to have some bearing on anxiety [8, 9], the mean max F_0 at the phonological-word (a content word plus zero or more postpositions) level for the same passage was 143 Hz (SD = 20) in the no-audience condition and 152 Hz (SD = 18) in the audience condition, the difference being highly significant, $t(130) = 8.51$, $p < .001$. These results are consistent with previous findings which suggest that speech-related anxiety is not only an effect of stuttering but also a cause [3]. But what is surprising is the magnitude of the effect size and the finding that even the presence of an audience consisting of one familiar, friendly person induces a great amount of stuttering.

Likewise, speech-related anticipatory anxiety is relevant to the questions given in the introduction, and most of the answers presented below were already suggested by previous researchers. Why does most stuttering occur in word- or sentence-initial position? When stutterers learn to stutter due to subtle speech-related anxiety, they soon come to mark the very first part of an utterance because that part is most conspicuous when they cannot initiate the utterance. In repeating this kind of experience, such sentence-initial sounds and words become fearful, some becoming feared sounds and words. This can also be a partial answer to the question why isolated words are stuttered less often than are sentences. While Salmelin et al. [10] hypothesize that "the lack of right-hemisphere activation in stutterers, time-locked to stimulus of speech onset, is associated with difficulties in initiating the correct prosody" (p. 1198), an alternative is that sentences are more fearful than isolated words because words are almost always embedded in sentences so that stutterers have had little opportunity to learn to stutter the same words in isolation.

Kalinowski and colleagues [11, 12] have recently attempted to account for the fact that in chorus reading, stuttering is dramatically reduced. They highlight the possible role of mirror neurons [13]. This hypothesis is neither verified nor refuted at present. We here present the following three interrelated or overlapping alternatives which seem equally plausible. First, in chorus reading, communicative responsibility is greatly decreased and the effect of the audience, which can be ignored, accordingly decreases [14]. This may also partially account for the reason why almost no stuttering occurs in singing, counting, and swearing [1]. Second, the stutterer's anxiety can be reduced because the effect of the pressure from the audience may be divided and reduced [3]. Third, in listening to the co-reader, an individual may shift attention away from the pressure coming from the audience and/or from the otherwise ensuing stuttering event, to the co-reader's reading performance. The attention shift may well explain the effects of DAF (delayed auditory feedback) and its variants. On the other hand, in repeated oral reading in the presence of an audience, the material and situation would become more familiar to the stutterer, and thus reduce anxiety.

The facts that there are four times more stutterers among males than females and that ~80% of developmental stuttering is spontaneously recovered may be related more or less to the behavior of subtle anticipatory anxiety. Boys are exposed more often to anxiety- and tension-evoking situations than are girls. But many stuttering children may sooner or later get accustomed to such situations, thus resulting in spontaneous recovery. If the main cause of stuttering is the presence of anxiety, stuttering would dissipate when anxiety disappears. This implies that spontaneous recovery takes place all of a sudden or in a short period of time, thereby making it difficult for researchers and parents to observe the moment of spontaneous recovery.

The answer to the final question, why words are easier to produce when first saying "ah" or using some other 'starter,' may be least associated with anxiety. That saying "ah" reduces the on-going anxiety effect does not seem to be very compelling. Instead, we should take this as just a simple fact [15]. In the case of the Japanese stutterer, "eh" /ee/, a common interjection in conversation, can be a good starter or reliever when the stutterer tenses the muscles of his speech organs in encountering feared sounds and words. This is because vowel (V) syllables are relatively easier to produce than consonant-vowel (CV) syllables. Also of interest in the case of the second author in an oral reading task in a no-audience condition is that the mean max F_0 of "eh" was about 108 Hz, which was close to an average inherent pitch level observed when male non-stutterers spontaneously produce this interjection in isolation. Thus this starter might function as a quick warm-up or springboard for the following utterance. Specifically, it may make utterances easier to produce by lowering the overall pitch level of the following utterance. We suggest that the use of this kind of starter constitutes part of the treatment of stuttering.

The treatment of stuttering

Given this line of discussion, we propose what we call the Overlearning Fluency when Alone (OFA) treatment consisting of two distinct but overlapping steps. Here, we are concerned only with oral reading, but the treatment can be applicable to spontaneous speech production. The first and most important step to take is to remove speech-related anticipatory anxiety, and this can easily be achieved when the stutterer is alone. The stutterer tries to read aloud an easy text while alone, and 'unlearns' stuttering by overlearning speech fluency in this carefree condition. As stated above, syllable repetitions do occur occasionally even in this condition, and it is here that we can capitalize on the function of the starter "eh," the insertion of which does not make oral reading unnatural. Indeed, the second author attempted to use the starter, "eh," in an oral reading task when he was alone. The result was almost complete disappearance of stuttering with no abnormality whatsoever.

So far, so good. But an important empirical question is if and how such fluency overlearned in a no-audience condition transfers to oral reading in the presence of an audience. Subtle speech-related anxiety precipitates stutterers to block fluent speech, and, what is worse, stutterers have overlearned such experiences. While the mechanism(s) involving the relationship between anxiety and stuttering is not well understood [16, 17], it seems difficult, albeit impossible, to 'unlearn' this naturally acquired, fully established relationship in ordinary treatment methods. Thus, the second step of our treatment is a kind of programmed (un)learning. Many years ago, Berwick [18] demonstrated that even the presence of a picture of a "hard listener" was enough to make stutterers uneasy and increase the frequency of stuttering. A present-day version of this is to use virtual reality environments, e.g., a virtual reality job interview environment [19]. The possibility of formulating a program learning approach for purposes of treatment where the audience size and audience trait are systematically varied is left to future research. On the other hand, research on the effect of audience in real situations where the audience consists of various types of people, such as the stutterer's mother, grandmother, siblings, and friends, is currently in progress in our project.

Conclusions

Speech-related anxiety, a main cause of stuttering, can easily and greatly be reduced in a no-audience condition, where one's true linguistic competence is best represented with little stuttering events accompanied. Overlearning of fluency in such an ideal condition, plus use of a natural starter in oral reading training, constitutes the first step of our OFA treatment. 'Unlearning' of the stuttering behavior which stutterers have acquired and reinforced for many years may appear extremely difficult, but the fact that their linguistic competence is unimpaired makes us optimistic. Our preliminary study suggests that it is not difficult to overlearn fluent, naturally-sounding speech in a no-audience condition. The next step is to try out the overlearned fluency in difficult situations. Depending upon how stable the foundation laid at the preceding step is, this step would become easier to take. We believe that we are now in good position to establish a simple and effective treatment for "one of mankind's most baffling afflictions."

REFERENCES

[1] Bloodstein, O. A Handbook on Stuttering, 5th Edn San Diego, CA: Singular; 1995.

[2] Buchel, O., Sommer, M. What causes stuttering? PLoS Biology 2004;2: 159-163.

[3] Johnson, W. Stuttering in children and adults. Minneapolis: University of Minnesota Press; 1955.

[4] Steer, M. D., Johnson, W. An objective study of the relationship between psychological factors and the severity of stuttering. J. Abnorm. Soc. Psychol 1936;31:36-46,

[5] Porter, H. V. K. Studies in the psychology of stuttering: XIV. Stuttering phenomena in relation to size and personnel of audience. J. Speech Dis 1939;4:323-333.

[6] Siegel, G. M., Haugen, D. Audience size and variations in stuttering behavior, J. Speech Hear. Res 1964;7:381-388.

[7] Armson, J., Foote, S., Witt, C., Kalinowski, J., Stuart, A. Effect of frequency altered feedback and audience size on stuttering. Eur. J. Disord. Com 1997;32:359-366.

[8] Hecker, M.H.L., Stevens, K.N., von Bismarck, G., Williams, C.E. Manifestations of task-induced stress in the acoustic speech signal. J. Acous. Soc. Amer 1968;44:993-1001.

[9] Scherer, K.R. Vocal affect expression: A review and a model for future research. Psychol. Bul 1986;99:143-165.

[10] Salmelin, R., Schnitzler, A., Schmitz, F., Freund, H. J. Single word reading in developmental stutterers and fluent speakers. Brain 2000;123: 1184-1202.

[11] Kalinowski, J., Dayalu, V.N. A common element in the immediate inducement of effortless, natural-sounding, fluent speech in people who stutter: 'the second speech signal,' Med. Hypotheses 2002;58:61-66.

[12] Kalinowski, J., Saltuklaroglu, T. Speaking with a mirror: engagement of mirror neurons via choral speech and its derivatives induces stuttering inhibition. Med. Hypotheses 2003;60:538-543.

[13] Rizzolatti, G., Arbib, M. Language within our grasp. Trends Neurosci 1998; 21:188-194.

[14] Eisenson, J., Wells, C. A study of the influence of communicative responsibility in a choral speech situation for stutterers. J. Speech Disord 1942;7:259-162.

[15] Swift, W.B., Hedrick, J. Sidetracking of stuttering by "starters." J. Applied Psychol 1917;1:84-88.

[16] Menzies, R. G., Onslow, M., Packman, A. Anxiety and stuttering: Exploring a complex relationship. Amr J Speech-Lang Pathol 1999;8:3-8.

[17] Craig, A., Tran, Y. Fear of speaking: chronic anxiety and stammering. Advan. Psychiatr. Treat 2006;12:63-68.

[18] Berwick, M. H. Stuttering in response to photographs of selected listeners, In Johnson, W. Stuttering in children and adults. Minneapolis: Unversity of Minnesota Press, pp. 274-277; 1955.

[19] Anderson, P.L., Zimand, E., Hodges, L.F., Rothbaum, B.O. Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure, Depress. Anxiety 2005;22:156-158.

Chapter 5

The audience/no-audience effects on stuttering, speech rate, and fundamental frequency in oral reading: A Japanese case

Co-authored with Takanobu Homma

Abstract

This case report investigated the audience/no-audience effects on stuttering, speech rate, and fundamental frequency (F_0) by analyzing the oral reading performance of an adult Japanese male stutterer in an A_1-B_1-A_2-B_2 paradigm, where A represents an audience phase and B a no-audience phase. The subject read a passage five times in each phase. The audience/no-audience effects on the three variables were found significant. The plausibility of F_0 as an indicator and 'precursor' of stuttering, in conjunction with the relation of anxiety to stuttering and F_0, is discussed. The possibility is also suggested that oral reading practice in a no-audience condition may be used as treatment for stuttering.

1. Introduction

Several studies have demonstrated that stutterers' speech is generally much more fluent when stutterers are alone than when they have an audience (Bergmann, 1987; Bloodstein, 1995; Hahn, 1940; Johnson, 1955; Porter, 1937; Yamada & Homma, 2007). We can draw at least two implications from these findings: (1) a major cause of stuttering is anxiety evoked in the presence of an audience (but cf. Messenger, Onslow, Packman, & Menzies, 2004) and (2) the true speech competence of stutterers, which is nearly normal, best manifests itself in the absence of an audience.

To learn more about these implications for clinical and research purposes, we investigated the oral reading performance of an adult Japanese stutterer in an A_1-B_1-A_2-B_2 paradigm, where A represents an audience-present phase and B a no-audience phase. The subject read the same passage five times in each phase, and we measured his reading performance in terms of stuttering frequency, speech rate, and fundamental frequency (F_0). On the basis of the findings from previous studies, we expected that the subject's reading performance would be much better in the no-audience condition, Phase B, displaying fewer stuttering events and faster speech rates.

Regarding fundamental frequency (F_0), we attempted to demonstrate that F_0 values tend to be lower in a no-audience condition than in an audience condition. It was more than half a century ago when Glasner and Dahl (1952), examining stuttering people, stated that "It has been interesting to observe that as improvement is noted, there is a definite drop in pitch level. The lowering of pitch seems to be one of the first indications of a lessening of tension" (p. 1113). But the relations between anxiety level, stuttering, and F_0 have not been well established in the literature (Atkinson, 1978; Bergmann, 1986; Hall & Yairi, 1992; Sacco & Metz, 1989; Scherer, 1979; Schmitt & Cooper, 1978; Yamada & Homma, 2007).

We also asked a question as to whether there is an adaptation effect (Bloodstein, 1995) in each phase, i.e., whether performance improves as the subject reads the same passage five times.

2. Method

2.1. Subject

The subject was the second author (T. H.) of this study, a 31-year-old

Japanese male speaker, who also served as the subjects in our preliminary study (Yamada & Homma, 2007). His stuttering, assessed as mild to moderate, appeared at the age of two. At the time of testing, he was a doctoral candidate and had a substantial knowledge of the mechanisms and treatments of stuttering.

2.2. Test material

The reading passage was taken from a book for general public written by Umesao (1988, p. 146), a Japanese naturalist. The text consisted of 117 *phonological words* and 476 syllables. A phonological word was defined as a constituent which consists of a content word and optionally following postpositions, e.g., "ima-de-wa," where "ima (now)" is a noun, "de" and "wa" are postpositions. (We will hereafter refer to a phonological word simply as a word.) Although the level of readability was unknown, our judgment was that it is at a high school level.

2.3. Procedure

The experiment used an A_1-B_1-A_2-B_2 design where Phase A represents an audience condition and Phase B a no-audience condition. The experimenter, the first author (J. Y.), a 55-year-old professor, with whom the subject was familiar, served as the only listener present. This condition was considered to be one of the situations which provoke little, if at all, anxiety in the subject.

The experiment was conducted in a small, relatively quiet room on the fifth floor of an eight-story university building which the subject used almost on a daily basis. The subject was seated at a small table opposite the experimenter. On the table were a Sony dynamic microphone F-V620, a Sony cassette-recorder TCM-1390, and a sheet of paper on which the reading material was printed.

Before the first trial of the audience condition, Phase A_1, the subject read silently the test passage until he was ready to read it orally. (No silent reading session was needed in the subsequent trials.) The experimenter then pressed the record button of the tape-recorder, and the first trial started. After the first trial was finished, a short break was given. In this manner, the subject orally read the passage five times in the first audience condition, Phase A_1. The experimenter then left the room, and the subject attempted five trials in the first no-audience condition, Phase B_1. On the first trial in this condition, the subject pressed the record button by himself, but he soon found himself somewhat tensed in doing so. He thus decided to keep the tape running after the second trial. He also thought that he would need a longer break to make himself more relaxed, and had longer breaks between the subsequent trials. On completing the fifth reading, he called the experimenter, and trials in the second audience condition, Phase A_2, began basically in the same manner as in Phase A_1. After completing five trials in this condition, the final five trials were given in the second no-audience condition, Phase B2, in a similar fashion.

2.4. Independent variables

The subject's oral reading performance in each condition was analyzed in terms of stuttering frequency, speech rate, and mean F_0 maximum value for words. Stuttering frequency was determined by the present authors, who independently assessed stuttered words listening to audiotape. Reliability was computed by dividing twice the stuttered words agreed upon by the total stuttered words identified by the two observers. The mean of the reliability coefficients for the 20 trials was 0.90 (SD = 0.05). Only the stuttering events agreed upon, i.e., clear cases, were used for main analysis.

Speech rate was defined in two ways. The first was the total reading time taken to read the passage. The second was mean duration for *fluent* words. This second measurement was considered necessary to ascertain whether slower speech rate, if indicated by total reading time, is largely due to higher stuttering frequency. A fluent word was defined as a word which neither of the authors identified as stuttered across the whole 20 reading trials. Of the 117 words in the test passage, 33 (28%) were fluent words, and the reading times for 660 such fluent words (33 words × 20 trials) were measured to the nearest millisecond using an Animo SUGI speech analyzer (Version 1.0.7.8).

The first author measured F_0 max values for a total of 2,340 words (117 words × 20 trials) using the same speech analyzer. To confirm the reliability of the measures, he quasi-randomly selected 10% of the total words and measured them again with an interval of more than one month. The correlation between the two measures was .998, $df = 232$, $p < .0001$, and the mean difference was .01 Hz.

3. Results

The main results for stuttering frequency (number of stuttering events), two speech rates, i.e., total reading time (sec) and mean durations of fluent words (msec), the mean F_0 max value (Hz) for each trial in each condition are presented in Table 1. In the following three sub-sections, a 2×2 ANOVA with repeated measures on the two factors, audience (audience and no-audience conditions, i.e., Phases A and B) and order (1st two phases and 2nd two phases), was performed using mean measurements for five trials in each phase as units for analysis.

[TABLE 1 GOES ABOUT HERE]

3.1. Stuttering frequency

The mean stuttering frequency was 42.8 (SD = 5.4), 19.4 (SD = 6.5), 33.6 (SD = 8.2), and 23.6 (SD = 4.9) in the first audience, first no-audience, second audience, and second no-audience conditions (Phases A_1, B_1, A_2, and B_2), respectively. The effect of audience was significant, $F(1, 4) = 25.67$, $p < .01$, but neither the effect of order nor the interaction between audience and order was significant, $F(1, 4) = 1.34$, n.s., and $F(1, 4) = 2.90$, $p > .1$, respectively. Paired comparisons indicated that the mean (42.8) in the first audience condition, Phase A_1, was significantly higher than those (19.4 and 23.6) in the first and second no-audience conditions, Phases B_1 and B_2, $t(4) = 4.94$, $p < .01$, and $t(4) = 6.04$, $p < .01$, respectively, and that the mean (33.6) in the second audience condition, Phase A_2, was significantly higher than that (19.4) in the first no-audience condition, Phase B_1, $t(4) = 3.20$, $p < .05$. These results showed that the subject stuttered significantly less often in the no-audience conditions than in the audience conditions. The mean (42.8) in the first audience condition, Phase A_1, tended to be higher than that (33.6) in the second audience condition, Phase A_2, $t(4) = 2.05$, $p < .11$, suggesting a positive effect of the preceding no-audience condition, Phase B_1, on the following audience condition, Phase A_2. The difference between the first and second no-audience conditions (19.4 and 23.6), Phases B_1 and B_2, and the difference between the second audience and second no-audience conditions (33.6 and 23.6), Phases A_2 and B_2, were not significant, $t(4) = 1.00$, and $t(4) = 1.84$, respectively. In summary, the results indicate (1) that, with less audience-associated anxiety in the no-audience conditions, Phases B_1 and B_2, the subject's oral reading performance was markedly improved and (2) that the preceding no-audience condition, Phase B_1, tended to improve the subject's performance in the following audience condition, Phase A_2.

3.2. Speech rate

The mean total reading time was 237 sec (SD = 31.6), 181 sec (SD = 47.5), 226 sec (SD = 51.1), and 192 sec (SD = 49.3) in the first audience, first no-audience, second audience, and second no-audience conditions (Phases A_1, B_1, A_2, and B_2), respectively. The effect of audience was significant, $F(1, 4) = 7.76$, $p < .05$, but neither the effect of order nor the interaction between audience and order was significant, $F(1, 4) < 1$. Paired comparisons indicated that the mean (237 sec) in the first audience condition, Phase A_1, was significantly longer than that (181 sec) in the first no-audience condition, Phases B_1, $t(4) = 2.36$, $p < .05$, and tended to be longer than that (192 sec) in the second no-audience condition, Phase B_2, $t(4) = 2.00$, $p < .06$. The other differences were not significant.

The mean duration for fluent words (N = 33) was 665 msec (SD = 16), 638 msec (SD = 17), 636 msec (SD = 10), and 626 msec (SD = 16) in the first audience, first no-audience, second audience, and second no-audience conditions (Phases A_1, B_1, A_2, and B_2), respectively. The pattern of results here was different from that for total reading time. Both the effect of audience and the effect of order were significant, $F(1, 4) = 8.05$, $p < .05$, and, $F(1, 4) = 16.11$, $p < .02$, but the interaction between audience and order was not, $F(1, 4) < 1$. Paired comparisons indicated that the mean (665 msec) in the first audience condition, Phase A_1, was significantly longer than that (638 msec) in the first no-audience condition, Phases B_1, $t(4) = 2.36$, $p < .05$, and tended to be longer than that (626 msec) in the second no-audience condition, Phase B_2, $t(4) = 2.00$, $p < .06$. The mean (665 msec) in the first audience condition, Phase A_1, was significantly longer than that (636 msec) in the second audience condition, Phase A_2, $t(4) = 3.05$, $p < .05$, and that (626 msec) in the second no-audience condition, $t(4) = 7.69$, $p < .01$, but the other differences were not significant.

This latter finding suggests that speech rate for fluent words may not be correlated with overall speech rate for the text. Indeed, as presented in Table 2, the correlation between mean duration for fluent words and total reading time, $r(18) = .44$, failed to reach the .05% level. The correlation between mean duration for fluent words and stuttering frequency, however, was significant, $r(18) = .61$, $p < .01$.

[TABLE 1 GOES ABOUT HERE]

3.3. Fundamental frequency

The mean F_0 max value was 152 Hz (SD = 1.8), 147 Hz (SD = 2.2), 148 Hz (SD = 2.0), and 149 Hz (SD = 2.6) in the first audience, first no-audience, second audience, and second no-audience conditions (Phases A_1, B_1, A_2, and B_2), respectively. Neither the effect of audience nor the effect of order reached significance, $F(1, 4) = 2.52$, $p > .1$, and $F(1, 4) = 5.12$, $p < .09$, but their interaction was significant, $F(1, 4) = 15.13$, $p < .02$. Paired comparisons showed that the mean (152 Hz) in the first audience condition, Phase A_1, was significantly higher than the mean (147 Hz) in the first non-audience condition, Phase B_1, the mean (148 Hz) in the second audience condition, Phase A_2, and the mean (149 Hz) in the second no-audience condition, Phase B_2, $t(4) = 7.07$, $p < .01$, $t(4) = 3.14$, $p < .05$, and $t(4) = 3.88$, $p < .05$, respectively. Interestingly, although the difference was small, the mean (147 Hz) in the first no-audience condition, Phase B_1, was significantly lower than that (149 Hz) in the second no-audience condition, Phase B_2, $t(4) = 5.72$, $p < .01$.

This pattern of results resembles those for stuttering frequency and mean duration for fluent words. As shown in Table 2, the correlation between stuttering frequency and F_0 was .63, $p < .01$, and that between mean duration for fluent words and F_0 was .67, $p < .001$.

3.4. Some more exploratory analysis of the relation between stuttering and F_0

It was worth exploring more about the relationship between stuttering and F_0. First, a point-biserial correlation between stuttering events (0 or 1) and F_0 values (df = 115) was carried out for each of the 20 trials. The mean for 20 correlation coefficients for the 20 trials was .29 (SD = .07), $p <.01$, ranging from .13 to .40. The correlations were significant for 18 out of the 20 trials, indicating that F_0 values were higher for stuttered words than for non-stuttered words. These findings, however, may be misleading because stuttering is more likely to occur at the beginning part of a sentence than toward the end, and because F_0 values are generally higher in sentence-initial position and decline toward the end of a sentence.

There would be at least two more ways to examine the relation between stuttering and F_0. The first is to compare F_0 for the stuttered and nonstuttered cases for the same words. We selected words which were stuttered 30% (6 trials) to 70% (14 trials) of the time in the 20 trials. There were 15 such words, and we compared mean F_0 values for their stuttered and nonstuttered cases. The mean was 156 Hz (SD = 12) for the stuttered group and 151 Hz (SD = 12) for the nonstuttered group, the difference being highly significant, $t(14) = 5.31$, $p < .001$. That is, F_0 was a mean of 5 Hz higher when the word was stuttered than when the same word was not stuttered.

The second way to reveal the relation between stuttering and F_0 is to compare F_0 values for stuttered words in a trial with a higher mean F_0 value on the one hand, and F_0 values for the same stuttered words in a trial with a lower mean F_0 value on the other. If we hypothesize that relatively higher F_0 values are associated with relatively more severely stuttered words, and further, if stuttered words are more severely stuttered in a trial with a higher mean F_0 value than in a trial with a lower mean F_0 value, then we would expect that F_0 values for stuttered words which appeared in a trial with a higher mean F_0 value would be higher than F_0 values for the same stuttered words which appeared in a trial with a lower mean F_0 value. To verify this hypothesis, we compared F_0 values for the same 10 stuttered words between the third trial of the first no-audience condition, Phase B_1, which exhibited the lowest mean F_0 condition, and the fourth trial of the first audience condition, Phase A_1, which exhibited the highest mean F_0 condition. (Note that one word which was stuttered in the former condition was not stuttered in the latter condition so that the number of the same words selected was not 11 but 10.) The result was what we expected. The mean for the former was significantly lower than that for the latter, 157 Hz (SD = 14) and 169 Hz (SD = 19), respectively, $t(9) = 5.36$, $p < .001$.

Given the findings above, we further asked a complementary question: what is the relationship between *nonstuttering* and F_0? A question we asked was whether F_0 values for the nonstuttered words which immediately precede stuttered words are higher than F_0 values for those which immediately precede nonstuttered words. Falck, Lawler, and Yonovitz (1985) gave a repeated-readings task to English-speaking stutterers (N = 7), and reported that the mean F_0 value tended to be *lower* for utterances prior to the moments of stuttering (pre-stuttered utterances) than that of identical utterances which were fluently produced (pre-nonstuttered utterances). Their finding was not consistent with our intuition, and thus we addressed the question of whether the same is true with a Japanese case. For our study, the question was whether nonstuttered words which immediately preceded stuttered words exhibit *lower* F_0 values than those which immediately precede nonstuttered words. We selected nine target words out of the 15 words which were stuttered 30% (6 trials) to 70% (14 trials) of the time in the 20 trials. Four words were excluded from the 15 words because they were in sentence-initial and/or pre-pause position, and two were excluded because less than five cases out of the 20 were nonstuttered.

Results showed that the mean F_0 was 152 Hz (SD = 19.6) for the pre-stuttered words and also 152 Hz (SD = 19.9) for the pre-nonstuttered words, $t(8) < 1$. Inspection of F_0 values for individual words, however, revealed that three nonstuttered words exhibited relatively large differences ($d > 5$ Hz) between the pre-stuttered and pre-nonstuttered conditions. Actually, a significant difference was observed for two of the words, one yielding 161 Hz (SD = 4.2) and 156 Hz (SD = 3.9), $t(17) = 2.45$, $p < .05$, and the other 164 Hz (SD = 5.5) and 157 Hz (SD =6.7), $t(16) = 2.45$, $p < .05$, in the pre-stuttered condition and in the pre-nonstuttered condition, respectively. The remaining word exhibited the opposite direction, but failed to reach significance, 163 Hz (SD = 13.0) and 170 Hz (SD = 5.2), N = 20, $t(12) = 1.6$, $p > .1$. Overall, these results were against the generality of Falck et al.'s finding.

Likewise, F_0 values for the nonstuttered words which immediately follow stuttered words were compared with F_0 values for those which immediately follow fluent words. There were 11 words which were appropriate for analysis, with the remaining four words excluded because they appeared in post-pause position and/or because the number of nonstuttered words was too small. The mean F_0 was 155 Hz (SD = 15.3) for the post-stuttered words and 154 Hz (SD = 13.6) for the post-nonstuttered words, $t(10) < 1$. There was one word for which the mean F_0 was significantly *lower* in the post-stuttered condition than in the post-nonstuttered condition, i.e., 124 Hz (SD = 7.8) and 131 Hz (SD = 2.9), N = 19, $t(13) = 2.47$, $p < .05$.

In sum, F_0 is closely associated with stuttering, but F_0 values of stuttered words are unlikely affect those of the immediately preceding or following nonstuttered words.

3.5. The adaptation effect

By inspection of Table 1, no apparent adaptation trend may be observed for any of the four variables (stuttering frequency, total reading time, duration for nonstuttered words, and F_0) in each phase. The speech rate for nonstuttered words appears to become much faster from the first audience condition, Phase A_1, to the first no-audience condition, Phase B_1, and F_0 also seems to become lower in a similar manner.

4. Discussion

This study observed several interesting results, and perhaps the most important ones involve these two: (1) audience/no-audience effects on stuttering frequency and speech rate in an A_1-B_1-A_2-B_2 paradigm, and (2) a possible relationship between anxiety, stuttering, and fundamental frequency F_0.

4.1. Stuttering frequency and speech rate

The association between anxiety and stuttering has been well documented for older children, adolescents, and adults (e.g., Bloodstein, 1995; Ezrati-Vinacour & Levin, 2004; Davis, Shisca, & Howell, 2006; Johnson, 1955;

but cf. Miller & Watson, 1992), but regarding the causal direction, Messenger et al. (2004) pointed out the possibility that "social anxiety is an effect of stuttering" (p. 207). The present study indicates that such is not likely to be the case; rather, stuttering is an effect of social anxiety. In addition to this basic finding, we will here discuss three intriguing yet under-researched characteristics of the audience/non-audience effects: audience type, fluctuations of performance, and the possibility for treatment.

The first is that even the presence of a *friendly* listener causes a great deal of stuttering, thus the absence of the listener markedly decreasing frequency of stuttering. To our knowledge, no previous study except for our pilot study (Yamada & Homma, 2007) has showen such a large effect of such a listener as an audience. In the present study, the number of stutterings was on average reduced to less than one half in the first no-audience condition, Phase B_1, as compared to the first audience condition, Phase A_1. This finding indicates that the present friendly listener failed to create a care-free, less anxiety-evoking situation. This failure then raises a question as to what criteria involving age, sex, trait, and others must be met to be a truly friendly, least-anxiety-inducing, ideal audience. In future research we would need to examine the effects of different types of audiences including virtual audience (e.g., Anderson, Zimand, Hodges et al. 2005; Brundage, Graap, Gibbons et al. 2006) and construct an audience-associated anxiety scale where the most demanding or threatening audience would be placed in the leftmost position, and a no-audience condition in the rightmost position. For clinical and research purposes, we would also need to examine more about the effects of the rightmost position on the scale, for fluctuations in stuttering frequency there were unexpectedly great.

Great fluctuations in the no-audience condition are taken as the second characteristic of the no-audience effect. Such fluctuations suggest that, even with no audience present, there occur some subtle external and/or internal stimuli which make him anxious, tense, or upset, resulting in more frequent stuttering events. As mentioned in the Method section, the subject felt uneasy when he pressed the record button in the first no-audience condition, Phase B_1. Also in the second no-audience condition, Phase B_2, he heard a chime ringing and footsteps outside of the room which he thought were disturbing his reading performance. Anecdotal stories abound about similar events and situations; for example, according to Bloodstein (1995), some stutterers say that "they may stutter if they hear a footstep down the hall, or if they talk 'as though' to a listener, or even if they think, 'If somebody were listening to me I would be stuttering'" (pp. 303-2).

New questions thus arise. What is the ideal no-audience condition like? Under an ideal condition, can stuttering frequency be kept at nearly zero? This latter question may probably be answered affirmatively, but the reality is that it is extremely difficult to have such an ideal condition. Anxiety-inducing events easily come and go, badly affecting a subject's oral reading performance on a moment-by-moment basis. The finding that the performance was somewhat poorer in the second no-audience condition, B_1, than in the first no-condition B_2, should be attributed to more occurrences of external/internal stimuli such as a ringing chime.

The third characteristic of the audience/no-audience effect is the tendency of the subject's performance to improve in the second audience condition, Phase A_2, as compared to that in the first audience condition, Phase A_1. This tendency is, although quite preliminary, encouraging. We emphasize that the subject exhibits nearly normal oral reading performance in anxiety-free situations, which represents his true competence. Thus, the critical question is how good performance in a no-audience condition can be transferred to audience-present situations. Specifically, how can we remove potential anxiety-causing stimuli in audience-present situations which precipitate him into stuttering. Aside from inherent difficulty identifying such anxiety-causing stimuli, one possibility would be for the subject to have a longer reading practice in a no-audience condition until he/she gains full confidence in the performance.

With regard to speech rate, we only note one point. That is that speech rate measured in terms of total reading time is qualitatively different from speech rate measured by mean duration for fluent words. This conclusion is suggested from the results that there was no significant correlation between them (Table 2), and that total reading time fluctuates more greatly from trial to trial than does mean duration for fluent words (Table 1). Overall fluency would better be represented by total reading time than by mean duration for fluent words. The higher correlation between stuttering and total time than between stuttering and mean duration for fluent words (Table 2) is taken as supporting evidence for that.

4.2. The relation between stuttering and F_0

Although the relationship between stuttering and F_0 has remained indeterminate in previous studies (Atkinson, 1978; Bergmann, 1986; Hall & Yairi, 1992; Sacc & Metz, 1989; Scherer, 1979; Schmitt & Cooper, 1978), the present results are clear-cut. As shown in Table 2, stuttering frequency significantly correlates with F_0. What remains to be done is to interpret the result. While the simple correlation between stuttering and F_0 allows several interpretations, we take it as suggesting that both stuttering and F_0 are an indicator of level of anxiety. The finding that mean F_0 became lower after the first audience condition, Phase A_1, is consistent with the notion that the higher state anxiety is ameliorated after the first audience condition. In retrospect, the subject thinks that this seems to have been the case.

The relation between anxiety and stuttering and F_0 was not straightforward, however. Stuttering and F_0 may differentially reflect level of anxiety. We will discuss the relation by considering two levels of effect: a global (or across-conditions) level where generalized state anxiety evoked by the experimental situation may be represented, and a local (or within-condition) level where specific state anxiety arising from particular kinds of stimuli may be represented. At a global level, a rough rule is that for F_0 to be lower than 150 Hz is necessary (though not sufficient) for stuttering frequency to be fewer than 30 per trial (Table 1). There was one exception to this rule (the fifth trial in the second no-audience condition, Phase B2), where the number of stutterings was 27 although the mean F_0 was 152 Hz. (Why this exception was observed is unknown. That the trial was the last one of the experiment might have something to do with it.) On the other hand, there were two trials (the third and fourth trials in the second no-audience condition, Phase B2) in which F_0 values were 149 Hz and 147 Hz whereas stuttering frequency was 34 and 41, respectively. For these trials, the effects exerting their influence at a local level might have caused more stutterings. For example, the words which the subject stuttered in the preceding audience condition might have become feared words, which in turn induced state anxiety and caused stutterings. Apart from some exceptions, therefore, at a global level, anxiety may first manifest itself as higher F_0 and then as stuttering. In other words, higher F_0 may be a 'precursor' to stuttering rather than the other way around.

At a local level, on the other hand, F_0 is likely to be higher when the word is stuttered than when it is not stuttered. It was noted, however, that there are little or no anticipatory and carry-over effects in the immediate vicinity of stuttered words. This implies that on a local level the rising or falling of F_0 is instantaneous, probably because anxiety appears or disappears instantaneously on this level.

4.3. Further questions

Several more questions are raised and we will briefly mention three of them. First, this was a case study, where the subject was the second author and the experimenter was the first author. Given this rather rare case, the generality of the present findings is unknown, and to what extent the present findings are generalized to other stutterers remains a challenge.

Second, can we use an A_1-B_1-A_2-B_2 paradigm for purposes of treatment for stuttering? The answer seems to be in the affirmative. For example, substantial practice in Phase B_1 may well result in good performance in the subsequent situation at least at a global level. Unfortunately, we failed to elicit the best oral reading performance that displays virtually no stuttering in the no-audience condition.

Finally, we conducted no psycholinguistic analysis in this study, and we ask whether such analysis in an A_1-B_1-A_2-B_2 paradigm may shed new light on treatment for individual stutterers. For example, feared and/or difficult sounds, if any, may profitably be compared between the audience and no-audience conditions. In the best trial in this study (the third trial of the first no-audience condition, B_1), for example, the subject produced 11 stutterings, where only five types of consonants, /s/, /t/, /ʃ/, /k/, and /d/, which may be identified as feared and/or difficult consonants, appeared. On the other hand, in a trial where many more stutterings were yielded, nearly all Japanese consonants were observed. What hierarchy of feared and/or difficult consonants exists remains another interesting question.

Acknowledgments

This study was supported in part by a Japan Ministry of Education, Science, and Technology Grant (No. 18530517). The authors are greatful Joe Lauer for his comments about an earlier draft of this paper.

References

Anderson, P. L., Zimand, E., Hodges, L. F., and Rothbaum, B. O. (2005). Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and Anxiety*, 22, 156-158

Atkinson, J. E. (1978). Correlation analysis of the physiological factors controlling fundamental voice frequency. *Journal of Acoustical Society of America*, 63, 211-222.

Bergmann, G. (1986). Studies in stuttering as a prosodic disturbance. *Journal of Speech and Hearing Research*, 29, 290-300.

Bloodstein, O. (1995). *A handbook on stuttering*. New York: Chapman & Hall.

Brundage, S. B., Graap, K., Gibbons, K. F. et al. (2006). Frequency of stuttering during challenging and supportive virtual reality job interviews. *Journal of Fluency Disorders*, 31, 325-339.

Erzrati-Vinacour, R. and Levin, I. (2004). The relationship between anxiety and stuttering: A multidimensional approach. *Journal of Fluency Disorders*, 29, 135-148.

Falck, F. J., Lawler, P. S., and Yonovitz, A. (1985). Effects of stuttering on fundamental frequency. *Journal of Fluency Disorders*, 10, 123-135.

Glasner, P. J. and Dahl, M. F. (1952). Stuttering – a prophylactic program for its control. *American Journal of Public Health*, 42, 1111-1115.

Hahn, E. F. (1940). A study of the relationship between the social complexity of the oral situation and the severity of stuttering. *Journal of Speech Disorders*, 5, 5-14.

Hall, K. D. and Yairi, E. (1992). Fundamental frequency, jitter, ans shimmer in preschoolers who stutter. *Journal of Speech and Hearing Research*, 35, 1002-1008.

Healey, E. C. and Gutkin, B. (1984). Analysis of stutterers' voice onset time and fundamental frequency contours during fluency. *Journal of Speech and Hearing Research*, 27, 219-225.

Homzie, M. J. and Lindsay, J. S. (1984). Language and the young stutterer: A new look at old theories and findings. *Brain and Language*, 22, 232-252.

Johnson, W. (1955). *Stuttering in children and adults: Thirty years of research at the University of Iowa*. Minneapolis: University of Minnesota Press.

Messenger, M. Onslow, M., Packman, A., and Menzies, R. (2004). Social anxiety in stuttering: measuring negative social expectancies. *Journal of Fluency Disorders*, 29, 201-212.

Miller, S. and Watson, B. (1992). The relationship between communication attitude, anxiety, and depression in stutterers and non-stutterers. *Journal of Speech and Hearing Research*, 35, 789-798.

Porter, H. V. K. (1939). Studies in the psychology of stuttering, XIV. Stuttering phenomenon in relation to size and personnel of audience. *Journal of Speech Disorders*, 4, 323-333.

Sacco, P. R. and Metz, D. E. (1989). Comparison of period-by-period fundamental frequency of stutters and nonstutterers over repeated utterances. *Journal of Speech and Hearing Research*, 32, 439-444.

Scherer, K. R. (1979). Personality markers in speech. In K. R. Scherer and H. Giles (eds.). *Social markers in speech*. Cambridge: CUP, pp. 147-209.

Schmitt, L. S. and Cooper, E. B. (1978). Fundamental frequencies in the oral reading behavior in stuttering and nonstuttering children. *Journal of Communication Disorders*, 11, 17-23.

Watson, B. and Alfonso, P. (1987). Physiological bases of acoustic LRT in nonstutterers, mild stutterers, and severe stutterers. *Journal of Speech* *and Hearing Research*, 30, 434-447.

Umesao, T. (1988). *Nihongo to nihon bunmei [The Japanese language and Japanese civilization].* Tokyo: Kumon Shuppan.

Yamada, J. and Homma, T. (2007). A simple and effective treatment for stuttering: Speech practice without audience. *Medical Hypothesis*, doi:10.1016/j.mehy.2007.02.045.

Table 1

Mean of SFr, stuttering frequency, TRT, total reading time (sec), DFW, mean duration for fluent words (msec), and MF_0, mean F_0 (Hz), in Each Trial

Trial	1	2	3	4	5	Mean
The first audience condition, Phase A_1						
SFr	36	46	42	50	40	42.8
TRT	224	211	230	229	292	237
DFW	646	653	677	682	671	665
MF_0	150	152	151	155	153	152
The first no-audience condition, Phase B_1						
SFr	27	17	11	17	25	19.4
TRT	259	154	144	154	192	181
DFW	656	641	640	620	663	638
MF_0	144	147	148	148	150	147
The second audience condition, Phase A_2						
SFr	42	26	34	41	25	33.6
TRT	259	191	202	299	179	226
DFW	646	639	638	636	620	636
MF_0	150	147	149	147	145	148
The second no-audience condition, Phase B_2						
SFr	18	25	29	19	27	23.6
TRT	155	163	272	163	209	192
DFW	601	634	631	644	632	626
MF_0	145	148	150	149	152	149

Table 2

Correlations between SFr, stuttering frequency, TRT,

total reading time (sec), DFW, mean duration for fluent words (msec), and

MF_0, mean F_0 (Hz), for the 20 Trials (df = 18)

	SFr	TRT	DFW	MF_0
SFr	1			
TRT	.73**	1		
DFW	.61*	.44	1	
MF_0	.63*	.33	.67**	1

*p < .01, **p < .001.